

ACTA DE EVALUACIÓN DE LA TESIS DOCTORAL
(FOR EVALUATION OF THE ACT DOCTORAL THESIS)

Año académico (academic year): 2017/18

DOCTORANDO (candidate PHD): **MARTÍN MONCUNILL, DAVID**

D.N.I./PASAPORTE (Id.Passport): ******7915R**

PROGRAMA DE DOCTORADO (Academic Committee of the Programme): **D442-INGENIERÍA DE LA INFORMACIÓN Y DEL CONOCIMIENTO**

DPTO. COORDINADOR DEL PROGRAMA (Department): **CIENCIAS DE LA COMPUTACIÓN**

TITULACIÓN DE DOCTOR EN (Phd title): **DOCTOR/A POR LA UNIVERSIDAD DE ALCALÁ**

En el día de hoy 04/05/18, reunido el tribunal de evaluación, constituido por los miembros que suscriben el presente Acta, el aspirante defendió su Tesis Doctoral **con Mención Internacional** (In today assessment met the court, consisting of the members who signed this Act, the candidate defended his doctoral thesis with mention as International Doctorate), elaborada bajo la dirección de (prepared under the direction of) MIGUEL ÁNGEL SICILIA URBÁN // ELENA GARCÍA BARRIOCANAL.

Sobre el siguiente tema (Title of the doctoral thesis): **ADOPCIÓN Y USO DE SISTEMAS DE ORGANIZACIÓN DEL CONOCIMIENTO ESPECÍFICOS DE DOMINIO: ANÁLISIS Y EVALUACIÓN DE FACTORES LIMITANTES**

Finalizada la defensa y discusión de la tesis, el tribunal acordó otorgar la CALIFICACIÓN GLOBAL¹ de **(no apto, aprobado, notable y sobresaliente)** (After the defense and defense of the thesis, the court agreed to grant the GLOBAL RATING (fail, pass, good and excellent): **Sobresaliente**

Alcalá de Henares, a 4 de Mayo de 2017

Fdo. (Signed): 

Fdo. (Signed): 

Fdo. (Signed): 

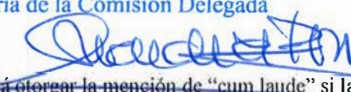
FIRMA DEL ALUMNO (candidate's signature),


Fdo. (Signed): DAVID MARTÍN MONCUNILL

Con fecha 11 de junio de 2018 la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado, a la vista de los votos emitidos de manera anónima por el tribunal que ha juzgado la tesis, resuelve:

- ☒ Conceder la Mención de "Cum Laude"
☐ No conceder la Mención de "Cum Laude"

La Secretaria de la Comisión Delegada



¹ La calificación podrá ser "no apto" "aprobado" "notable" y "sobresaliente". El tribunal podrá otorgar la mención de "cum laude" si la calificación global es de sobresaliente y se emite en tal sentido el voto secreto positivo por unanimidad. (The grade may be "fail" "pass" "good" or "excellent". The panel may confer the distinction of "cum laude" if the overall grade is "Excellent" and has been awarded unanimously as such after secret voting.)



Universidad
de Alcalá

COMISIÓN DE ESTUDIOS OFICIALES
DE POSGRADO Y DOCTORADO

En aplicación del art. 14.7 del RD. 99/2011 y el art. 14 del Reglamento de Elaboración, Autorización y Defensa de la Tesis Doctoral, la Comisión Delegada de la Comisión de Estudios Oficiales de Posgrado y Doctorado, en sesión pública de fecha 11 de junio, procedió al escrutinio de los votos emitidos por los miembros del tribunal de la tesis defendida por *MARTÍN MONCUNILL, DAVID*, el día 4 de mayo de 2018, titulada *ADOPCIÓN Y USO DE SISTEMAS DE ORGANIZACIÓN DEL CONOCIMIENTO ESPECÍFICOS DE DOMINIO: ANÁLISIS Y EVALUACIÓN DE FACTORES LIMITANTES*, para determinar, si a la misma, se le concede la mención "cum laude", arrojando como resultado el voto favorable de todos los miembros del tribunal.

Por lo tanto, la Comisión de Estudios Oficiales de Posgrado **resuelve otorgar** a dicha tesis la

MENCIÓN "CUM LAUDE"

Alcalá de Henares, 12 de junio de 2018

EL VICERRECTOR DE INVESTIGACIÓN Y TRANSFERENCIA



J. Javier de la Mata de la Mata

Copia por e-mail a:

Doctorando: MARTÍN MONCUNILL, DAVID

Secretario del Tribunal: JUAN JOSÉ CUADRADO GALLEGO

Director/a de Tesis: MIGUEL ÁNGEL SICILIA URBÁN // ELENA GARCÍA BARRIOCANAL



Universidad
de Alcalá

ESCUELA DE DOCTORADO
Servicio de Estudios Oficiales de
Posgrado

DILIGENCIA DE DEPÓSITO DE TESIS.

Comprobado que el expediente académico de D./D^a _____
reúne los requisitos exigidos para la presentación de la Tesis, de acuerdo a la normativa vigente, y habiendo
presentado la misma en formato: ☐ soporte electrónico ☐ impreso en papel, para el depósito de la
misma, en el Servicio de Estudios Oficiales de Posgrado, con el nº de páginas: _____ se procede, con
fecha de hoy a registrar el depósito de la tesis.

Alcalá de Henares a _____ de _____ de 20____



Fdo. El Funcionario

UNIVERSIDAD DE ALCALÁ
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN, ESCUELA POLITÉCNICA



DOCTORADO EN INGENIERÍA DE LA INFORMACIÓN Y DEL CONOCIMIENTO
TESIS DOCTORAL

ADOPCIÓN Y USO DE SISTEMAS DE ORGANIZACIÓN DEL
CONOCIMIENTO ESPECÍFICOS DE DOMINIO: ANÁLISIS Y EVALUACIÓN
DE FACTORES LIMITANTES

Autor:

David Martín Moncunill

Directores:

Dr. Miguel Ángel Sicilia Urbán

Doctor en Informática

Universidad de Alcalá

Dra. Elena García Barriocanal

Doctora en Informática

Universidad de Alcalá

Alcalá de Henares, 18 de Julio de 2017

UNIVERSITY OF ALCALÁ
COMPUTER SCIENCE DEPARTMENT, POLYTECHNIC SCHOOL



DOCTORADO EN INGENIERÍA DE LA INFORMACIÓN Y DEL CONOCIMIENTO
DOCTORAL THESIS

ADOPTION AND USE OF DOMAIN-SPECIFIC KNOWLEDGE
ORGANIZATION SYSTEMS: ANALYSIS AND EVALUATION OF LIMITING
FACTORS

Author:

David Martín Moncunill

Supervisors:

Dr. Miguel Ángel Sicilia Urbán

Doctor en Informática

Universidad de Alcalá

Dra. Elena García Barriocanal

Doctora en Informática

Universidad de Alcalá

Alcalá de Henares, 18th July 2017

Dr. D. José Javier Martínez Herráiz, Profesor Titular de Universidad, del Área de Ciencias de la Computación e Inteligencia Artificial, en calidad de Coordinador de la Comisión Académica del Programa de Doctorado Ingeniería de la Información y el Conocimiento,

CERTIFICO: Que la Tesis Doctoral titulada "Adopción y uso de sistemas de organización del conocimiento específicos de dominio: análisis y evaluación de factores limitantes" realizada por D. David Martín Moncunill, y dirigida por el Dr. D. Miguel Angel Sicilia Urbán y la Dra. Dña. Elena García Barriocanal, reúne los requisitos para su presentación y defensa pública.

En Alcalá de Henares, a 1 de diciembre de 2017

Fdo.: D. José Javier Martínez Herráiz





Universidad
de Alcalá

DPTO. DE CIENCIAS DE LA COMPUTACIÓN
Escuela Politécnica Superior
Campus Universitario, Edificio Politécnico
28871 Alcalá de Henares (Madrid)
Teléfonos: 91 885 6645 Fax: 91 885 6646
e-mail: dpto.ccomputacion@uah.es

Dra. Dña. Elena García Barriocanal, Profesora Catedrática de Universidad del Área de Lenguajes y Sistemas Informáticos del Departamento de Ciencias de la Computación de la Universidad de Alcalá,

Dr. D. Miguel Ángel Sicilia Urbán, Profesor Catedrático de Universidad del Área de Lenguajes y Sistemas Informáticos del Departamento de Ciencias de la Computación de la Universidad de Alcalá,

CERTIFICAN: Que una vez concluido el trabajo de Tesis Doctoral titulado "Adopción y uso de sistemas de organización del conocimiento específicos de dominio: análisis y evaluación de factores limitantes" realizado por D. David Martín Moncunill, dicho trabajo tiene suficientes méritos teóricos, que se han contrastado adecuadamente mediante validaciones experimentales y que son altamente novedosos. Por todo ello se considera que procede su defensa pública.

Y para que conste, firma la presente en Alcalá de Henares, a 22 de diciembre de 2017

El Director de la Tesis

La Directora de la Tesis

Fdo.: Dr. Miguel Ángel Sicilia Urbán

Fdo.: Elena García Barriocanal



***Here's to the ones who dream.
Foolish as they may seem.
Here's to the hearts that ache.
Here's to the mess we make.***

("The fools who dream", La La Land)

UNIVERSIDAD DE ALCALÁ
DEPARTAMENTO DE CIENCIAS DE LA COMPUTACIÓN, ESCUELA POLITÉCNICA



DOCTORADO EN INGENIERÍA DE LA INFORMACIÓN Y DEL CONOCIMIENTO
TESIS DOCTORAL

ADOPCIÓN Y USO DE SISTEMAS DE ORGANIZACIÓN DEL
CONOCIMIENTO ESPECÍFICOS DE DOMINIO: ANÁLISIS Y EVALUACIÓN
DE FACTORES LIMITANTES

Autor:

David Martín Moncunill

Directores:

Dr. Miguel Ángel Sicilia Urbán

Doctor en Informática

Universidad de Alcalá

Dra. Elena García Barriocanal

Doctora en Informática

Universidad de Alcalá

Agradecimientos

Fui uno de esos tantos aspirantes a ingeniero técnico que comenzó a trabajar bastante antes de terminar la carrera y por este y otros motivos tardé más de lo esperado en terminarla. Seguramente mucho más de lo que hubieran esperado los profesores que me acompañaron hasta mi ingreso en la universidad, de los cuales guardo un gran recuerdo y que sin duda contribuyeron para que a día de hoy esté escribiendo estas líneas. Tanto me costó terminar la carrera que tras presentar mi TFC, cuando se me ofreció unirme a un grupo de investigación de la universidad para trabajar al frente de proyectos europeos de investigación, no pude evitar que saliera mi orgullo, con un razonamiento similar al que sigue:

- “Era un buen estudiante y volveré a serlo. Voy a llegar hasta la máxima titulación académica”.

La realidad es que comencé el doctorado sin pensar en las consecuencias, en si vendría bien para mi carrera laboral o en el provecho que podría sacarle en contraste con el esfuerzo a invertir. Mentiría si no dijera que esa decisión se basó en el orgullo, en “quitarme la espinita” de no haber estado a la altura durante el estudio de la carrera y en demostrar lo que era capaz de realizar. He de agradecer por tanto a aquellas personas, que aún sin quererlo, hicieron aflorar ese orgullo que me hizo tomar la decisión que me ha llevado hasta aquí.

Me ha llevado 5 años, donde los dos primeros fueron para obtener el Grado en Sistemas de Información y posteriormente el Máster que me daba acceso al doctorado. Todo esto sin dejar de trabajar en proyectos de europeos (11 en total) realizando distintas tareas, la mayoría ligadas al “*management*”, a las pruebas con usuarios y a la diseminación y explotación. Tampoco han faltado colaboraciones docentes tanto en grado como máster, impartición de cursos de extensión universitaria y la creación de un par de MOOCs; resumiendo, he dormido bastante poco.

Evidentemente he aprendido mucho, pero lo que más agradezco son las capacidades que he reforzado durante estos años como la de adaptación, negociación, trabajo en entornos multidisciplinares y multiculturales, liderazgo y otras con las que tampoco cabe extenderse, ya que no se trata de un currículo.

Lo que sí cabe es agradecer a todas las personas que me han acompañado en el desarrollo de estos proyectos de investigación, comenzando por los directores de mi tesis, la Dra. Elena García Barriocanal y el Dr. Miguel Ángel Sicilia Urbán a quienes he procurado agradecerles el esfuerzo invertido en mi persona con una tesis que esté a la altura del mismo. También me gustaría agradecer la labor del Dr. Salvador Sánchez Alonso, con el que he trabajado codo con codo en multitud de proyectos de investigación y con quien he disfrutado de muchos viajes asociados a los mismos.

Igualmente debo agradecer el apoyo de mis compañeros de laboratorio del grupo de investigación IERU a quienes he capturado en la siguiente fotografía, destacando especialmente la figura del Dr. Paulo Alonso Gaona García, que me brindó la oportunidad de colaborar con mis conocimientos en usabilidad, apareciendo así como coautor en mi primer artículo JCR. En la foto falta Juan, otro buen amigo que decidió acompañarme en mi siguiente parada en Bureau Veritas, tras abandonar la universidad.



Mi familia siempre se ha interesado por mis estudios y mis padres siempre han hecho todo lo posible por facilitarme las cosas, aunque yo me haya negado en repetidas ocasiones a aceptar su ayuda; les agradezco enormemente su paciencia. A mi padre, mi madre, mi hermana, abuelos y toda mi familia les mando mi más sincero agradecimiento.

Fuera del entorno familiar he de agradecer especialmente a José Manuel y David Baños, mis amigos más cercanos (y compinches de www.geektopia.es), los que mejor me han podido comprender, como ingenieros que son (de teleco), azuzar cuando ha sido necesario e incluso cuando no lo era. Hablando de telecos, no querría olvidarme de mis amigos de la Asociación de Actividades Culturales (Kultu) y la Rama del IEEE de la escuela de “Teleko” de “Vallekas”. Sé que Richi “llorará” si no le menciono expresamente, así que lo haré y aprovecharé también para mencionar a Claudia, Emilio y Pablo.

Si ha habido algo que durante estos años me ha servido de vía de escape y me ha acompañado ha sido mi pasión por las artes marciales y, más concretamente, el Kenpo. Nunca he dejado de entrenar y al “mudarme” a la UAH quise montar clases, cosa que logré con gran éxito gracias al apoyo de Fernando Gil y el respaldo de Francisco Guzón.

Ahora que, por motivos laborales, me he vuelto a mudar a Barcelona, las clases siguen funcionando gracias a Alberto y Néstor, amigos y alumnos míos de cuando impartía clases de Kenpo en la Universidad Politécnica de Madrid; clases que actualmente también siguen en pie, dirigidas por Jesús y Carlos, igualmente amigos y alumnos. En la siguiente foto salimos todos junto con el Dr. Ing. Danilo Simón Zorita, sin lugar a dudas un gran apoyo durante todos estos años.



Del grupo de Kenpo de la UAH quiero agradecer especialmente el cariño de Alberto Mariscal Torres, futuro licenciado en CAFYDE, que tuvo a bien alojarme en su casa durante un buen periodo de tiempo. También a David Casillas, quién actualmente se está doctorando, a Inma que seguro será una excelente profesional farmacéutica, a Sergio que se encarga de revisar el nivel de dureza en las clases, a Ceci que incluso aportó en la revisión de esta tesis, al futuro doctor (en medicina) Jesús, a la doctora (también en medicina) Sandra, a los ya ingenieros informáticos David y Nuria, a César y Víctor (que empezaron la carrera conmigo), a María Elena y a todos mis alumnos, tanto de la UAH como de la UPM que me han cuidado tan bien durante estos años.



Evidentemente, si imparto clases de Kenpo es porque he tenido un maestro que me ha guiado en el camino, Sigung José Ignacio Vegue, al que le quiero agradecer su apoyo todos estos años al igual que a Sifu Carlos Sanz y a todos los compañeros que han tenido a bien colaborar en mi aprendizaje intercambiando golpes y cervezas: Luis, Javier Marcos, Raquel, Miguelón, César, Iván y tantos otros. Más de 21 años entrenando Kenpo da para conocer a bastante gente, pero entre la gente que destaca debo mencionar a Rafael Carriet que, junto a su familia (mis ánimos a Mayte que comienza su carrera universitaria) y alumnos, siempre me han acogido de excelente forma en mis visitas a Euskadi.

En cuanto a las visitas, tampoco debo olvidarme que he realizado dos estancias internacionales: gracias al Dr. Christian Stracke (Open University in the Netherlands) y al Dr. Carlos Enrique Montenegro (Universidad Distrital Francisco José de Caldas). Mi más sincero agradecimiento especialmente a la familia Montenegro que me acogió de manera espectacular, como si de uno más se tratara.

Son muchas las personas que me han acompañado durante todos estos años y es imposible seguir mencionando gente sin que estos agradecimientos se hagan grotescamente largos en comparación con la tesis. Finalizaré agradeciendo a la persona que me ha estado acompañando en la fase final de la escritura de la tesis. Moltes gràcies Sandra, a tu i a la Clàudia per cuidar-me tant aquests mesos a Barcelona.

Resumen

Los sistemas de organización del conocimiento (KOS – “*Knowledge Organization Systems*”) son una herramienta de gran ayuda de cara a la clasificación, organización y localización de información. La construcción de un KOS parte de una serie de términos, cuya selección en la mayor parte de los casos atiende a un dominio específico. Sobre estos términos se establecen distintas relaciones, como podría ser una relación jerárquica o una relación de equivalencia. La forma más elemental de uso de KOS consiste en valerse de los términos que los componen como palabras clave, mientras que las relaciones existentes permiten desarrollar funcionalidades avanzadas empleadas en áreas como la inteligencia artificial o la web semántica, entre otras. Los usuarios pueden beneficiarse de esta información de manera consciente mediante el empleo de interfaces que permiten operar con los términos y relaciones de un KOS, ya sea de manera textual o visual. La investigación relativa al desarrollo de estas interfaces tiene un largo recorrido y múltiples estudios muestran las potenciales ventajas que ofrecen para el usuario final. Sin embargo, a nivel práctico, su uso no está extendido en este sentido, contradiciendo los beneficios que resultan a nivel teórico. El presente trabajo persigue indagar en algunas de las posibles causas de este desuso y contribuir aportando metodologías que permitan evaluar su impacto, proporcionando información de utilidad para modificar un KOS de forma que se potencie su uso.

Palabras Clave

Sistemas de Organización del Conocimiento, especificidad de dominio, representación del conocimiento, usabilidad, recuperación de información, visualización de información, metadatos.

Abstract

Knowledge Organization Systems (KOS) are a helpful tool for the classification, organization and location of information. The construction of a KOS starts with the selection of a number of terms, which in most cases are related to a specific domain. Then different kinds of relationships are established linking those terms, for instance a hierarchical or an equivalence relation. The most elementary form of use of KOS is to make use of the terms that compose them as keywords while existing relationships enable advanced features which are commonly used in areas such as artificial intelligence and semantic web. Users can consciously benefit from this information through the use of interfaces that enable to operate with KOS terms and relationships, either in a textual or visual way. Research on the development of these interfaces has a long history and multiple studies have shown the advantages for the end user. However, after several years of research aiming to improve the use of KOS through different interfaces, they are not widespread, contradicting the benefits resulting theoretical level and in controlled experiments. This thesis aims to investigate some of the possible causes of this disuse and provide methodologies to evaluate the impact of these causes, aiming to obtain useful information to promote the use of KOS.

Keywords

Knowledge Organization Systems, domain specificity, knowledge representation, usability, information retrieval, classification, information visualization, metadata.

Índice General

Agradecimientos	I
Resumen	V
Abstract.....	VI
Índice General.....	XX
Índice de Figuras	XXII
Índice de Tablas.....	XXIV
Capítulo 1: Introducción	1
1.1 Objeto de estudio	1
1.2 Planteamiento del problema	2
1.4 Metodología de trabajo	9
1.5 Estructura del documento	11
Capítulo 2: Estado de la cuestión.....	12
2.1 Introducción sobre Sistemas de Organización del Conocimiento.....	12
2.2 Principales tipos y usos de KOS	14
2.2 KOS de dominio específico	20
2.3 Distancia de especialización entre conceptos.....	23
2.4 Extracción automática de palabras clave.....	26
2.5 Interfaces y visualización de KOS: tendencias y desafíos.	31
Capítulo 3: Mecanismos de evaluación y análisis de los factores limitantes observados	37
3.1 Introducción	37
3.2 Evaluación de la aplicabilidad práctica de la extracción automática de palabras clave	39
3.3 Evaluación del grado de especificidad de dominio en los términos de un KOS.....	55
3.4 Evaluación de la distancia de especialización entre conceptos.....	73
Capítulo 4: Resumen de los experimentos realizados en cuanto a interfaces visuales con KOS	87
4.1 Motivación	87
4.2 Introducción	87
4.3 <i>Applying visualization techniques to develop interfaces for educational repositories: the case of Organic.Lingua and VOA3R</i>	89
4.4 <i>A usability study of taxonomy-visualization user interfaces in digital repositories</i>	91
4.5 <i>Selection and Use of Search Mechanisms in Learning Object Repositories: the Case of Organic.Edunet</i>	94
4.6 <i>Estudios publicados relacionados con interfaces que emplean realidad virtual</i>	99
Capítulo 5: Conclusiones	103
5.1 Consecución de objetivos	105
5.2 Respuesta a las preguntas de investigación	108

5.3	Aportaciones.....	112
5.4	Trabajo futuro y conclusión final.....	118
Referencias Bibliográficas.....		122
Anexo I: Interfaces existentes en Organic.Edunet		134
Anexo II: Guía de aplicación de la metodología para la evaluación de la extracción automática de palabras clave		137
Anexo III: Guía de aplicación de la metodología para la evaluación de la especificidad de dominio de un KOS		140
Anexo IV: Guía de aplicación de la metodología para la asignación y evaluación de la distancia de especialización entre conceptos		145

Índice de Figuras

Figura 1. Estructuras y estrategias para organización del conocimiento. Extraída de (Gaona-García P. A., Visualización de esquemas de representación de conocimiento para el acceso a recursos en repositorios digitales, 2014) y basada en (Lei, 2008).....	17
Figura 2. Razonamientos que pueden surgir al navegar por un tesoro en un contexto educativo en base a los experimentos realizados.	19
Figura 3. Recursos de AGRIS por año hasta 2013 (extraído en Noviembre de 2013).....	43
Figura 4. Recursos de AGRIS por tipo de recurso hasta 2013 (extraído en Noviembre de 2013)	43
Figura 5. Un fragmento de los metadatos de un recurso descritos siguiendo el perfil de aplicación de AGRIS (AGRI AP).....	44
Figura 6. Distribución de palabras clave en inglés de la colección VOA3R-AGRI45	45
Figura 7. Detalle del número de recursos de la colección VOA3R-AGRI según su número de palabras clave, en el intervalo de 1 a 10 palabras clave.	46
Figura 8. La base de datos relacional creada, mostrando información sobre los recursos provenientes de los Países Bajos (NL) en el año 2000.	47
Figura 9. Información contenida en un número de registro ARN	47
Figura 10. Número de términos clasificados por su profundidad para cada uno de los 4 grupos establecidos en función del tamaño asignado a la rama.	60
Figura 11. Resultados de especificidad de los términos en porcentaje, por ramas.....	64
Figura 12. Resultados de especificidad de los términos en porcentaje (ordenadas) en relación al nivel de profundidad (abscisas).	65
Figura 13. Todos los términos de este conjunto, proveniente de la rama “Technology” son genéricos, los términos con profundidad “5” y “6” están subrayados en rojo.	66
Figura 14. Términos coincidentes con las ontologías SUMO y MILO (escala de la derecha - 0,200) con los extraídos en la evaluación realizada por expertos humanos (escala de la izquierda - 0,900).....	71
Figura 15. Escenario basado en una interfaz textual sin emplear la información de distancia de especialización entre conceptos.	80
Figura 16. Escenario basado en una interfaz textual empleando la información de distancia de especialización entre conceptos.	80
Figura 17. Escenario basado en una interfaz visual sin emplear la información de distancia de especialización entre conceptos.	81
Figura 18. Escenario basado en una interfaz visual empleando la información de distancia de especialización entre conceptos.	81
Figura 19. Interfaz navegacional de Organic.Edunet.....	89
Figura 20. Interfaz navegacional de VOA3R	90
Figura 21. Interfaz navegacional de VOA3R	90
Figura 22. Las ocho interfaces visuales desarrolladas para el estudio	91
Figura 23. Una de las interfaces empleadas para la primera parte del estudio. El tipo de visualización de la interfaz es radial, el tamaño de los círculos representa la cobertura.....	92

Figura 24. Una de las interfaces empleadas para la segunda parte del estudio, denominada como “Sunburst”, también de tipo radial, el tamaño de los sectores representa la cobertura. ...	93
Figura 25. Ocasiones en las que se intentó emplear una interfaz para completar una tarea y su tasa de éxito.	96
Figura 26. Tasa de éxito para cada interfaz de búsqueda en la tarea C.	97
Figura 27. Tiempos y errores en la tarea C para cada usuario (las columnas con relleno punteado indican que la tarea no se completó satisfactoriamente).	97
Figura 28. Recomendaciones de desarrollo para la interfaz de Organic.Lingua.	98
Figura 29. Representación simple del KOS en entorno 3D, centrado en los animales presentes en los humedales.	100
Figura 30. Una visión más amplia del KOS en Unity SDK.	101
Figura 31. Representación de los humedales en el ambiente 3D que recrea la zona.	101
Figura 32. Representación de uno de los animales en el ambiente 3D.	102
Figura 33. Utilidad, usabilidad, deseo de uso y experiencia de usuario.	119

Índice de Tablas

Tabla 1. Relación que presentan las tácticas de búsqueda de Bates con la relación "gen-spec", según García y Sicilia (2003)	24
Tabla 2. Resultados para la evaluación automática de la indexación, en términos de precisión y exhaustividad utilizando la muestra seleccionada de 2000 documentos (el corpus de entrenamiento tenía 1200 documentos).....	49
Tabla 3. Resultados calculados para la evaluación automática de la indexación, en términos de precisión, exhaustividad y medida F. Derivado de la Tabla 1 en Lim et al. (2013).....	50
Tabla 4. Resultados obtenidos por los expertos humanos en términos de precisión, exhaustividad medida F.	50
Tabla 5. Primeras filas de los resultados reportados por el evaluador A.	51
Tabla 6. Resultados que cada evaluador humano asignó a las palabras clave extraídas por KEA	51
Tabla 7. Resultados globales de la evaluación manual.	52
Tabla 8. Análisis de la varianza.....	52
Tabla 9. Los 25 conceptos raíz de AGROVOC: número de términos en cada rama, nivel de profundidad y tamaño asignado en función a los mismos.	59
Tabla 10. Conclusiones de la evaluación interna para la selección de los términos.	61
Tabla 11. Ejemplo abreviado del cuestionario con el que se obtuvieron los datos, el fragmento mostrado corresponde a un caso real.	62
Tabla 12. Valores asignados por expertos, media calculada y categorización de especificidad de dominio para algunos de los términos analizados	64
Tabla 13. Casos inciertos y motivación detectada para cada uno de ellos tras el análisis cualitativo.....	67
Tabla 14. Resultados de especificidad obtenidos tras la aplicación del criterio propuesto.	69
Tabla 15. Resultados de contrastar el set de términos elaborado en base a la evaluación de los expertos humanos con los términos existentes en las ontologías SUMO y MILO.	70
Tabla 16. BT de "Foods"	77
Tabla 17. Formulario mostrando la jerarquía completa (extracto de 11 términos)	78
Tabla 18. Formulario mostrando solo los términos situados por debajo de Foods → Beverages (extracto de 11 términos)	79
Tabla 19. Distancia calculada para los términos situados por debajo de Foods → Beverages (extracto de los 11 términos mostrados en la tabla 18)	80
Tabla 20. Extracto de la tabla de resultados para los términos de nivel de profundidad 3.	82
Tabla 21. Resultados globales del set de datos de la distancia de especialización entre conceptos.	83
Tabla 22. Número y porcentaje de participantes que prefirieron las interfaces que empleaban la distancia de especialización entre conceptos.....	84
Tabla 23. Tareas de búsqueda "C", "E" y "H" y métodos más adecuados para llevarlas a cabo.	95

Tabla 24. Resultados para la tarea “C.....	95
Tabla 25. Preguntas de investigación y objetivos asociados	104
Tabla 26. Consecución de los objetivos previstos.	105

Capítulo 1: Introducción

“Después de tantas aventuras que te he hecho correr junto a mí, después de tanta empresa loca equivocada, después de los molinos que tomé por gigantes y los cueros de vino que hice trizas, después de la Trifaldi y Clavileño, de la venta que tomé por castillo, de los galeotes que dimos en salvar, véome en el caso de cumplir la promesa que te hice, y hete aquí que te entrego, para tu gobierno, para tu protección y tu agonía, sin que puedas negarte o abandonarla, sin que puedas jamás dar la espalda a sus lides, en la certeza que sabrás guardarla, la ínsula que soy.”

(Extracto de “Don Quijote de la Mancha”, de Miguel de Cervantes, 1615)

1.1 Objeto de estudio

Los sistemas de organización del conocimiento (KOS – “Knowledge Organization Systems”) incluyen todos los tipos de esquemas para la organización del conocimiento como las taxonomías, los tesauros o las ontologías. La construcción de los diferentes tipos de sistemas de organización del conocimiento, parten de una serie de términos, sobre los que se establecen distintas relaciones (Kim & Cavedon, 2011), como podría ser una relación jerárquica o una relación de equivalencia.

Estos sistemas han demostrado su utilidad como herramienta para la clasificación, organización y localización de información (Lancaster, 1972) y aunque existen iniciativas para construir sistemas de organización del conocimiento generalistas como SUMO (Niles & Pease, 2001), MILO (Niles & Terry, 2004) o CYC (Lenat & Guha, 1990), los términos sobre los que se construyen los sistemas de organización del conocimiento suelen centrarse en un dominio (Mai, 2005) concreto.

Sobre los sistemas de organización del conocimiento se han desarrollado tecnologías que permiten crear, gestionar y visualizar modelos que muestren una perspectiva simplificada de los conceptos que forman un determinado dominio y la estructura semántica subyacente (Sánchez-Cuadrado & Morato, 2009); facilitando la comprensión, gestión y recuperación de los conceptos contenidos en cada dominio por parte del usuario.

En el ámbito de las tecnologías de la información se lleva realizando desde hace años un importante esfuerzo para explotar las posibilidades de uso de los sistemas de organización del conocimiento en dominios como la biomedicina (Kashyap, Morales, & Hongsermeier, 2006), la educación (Rius, Sicilia, & Barriocanal, 2007) o la gestión de información (Colomo-Palacios, García-Crespo, Soto-Acosta, Ruano-Mayoral, & Jiménez-López, 2010), entre otros muchos.

Una de las formas más elementales de uso de los sistemas de organización del conocimiento consiste en valerse de los términos que los componen, empleándolos como palabras clave, de manera que sirvan para clasificar la información, permitiendo así acciones como acotar resultados en búsquedas mostrando únicamente los recursos identificados con una serie de términos. Por otro lado, las relaciones que se establecen entre los términos de los sistemas de organización del conocimiento permiten desarrollar funcionalidades avanzadas de utilidad en áreas como la inteligencia artificial o la web semántica.

Los usuarios pueden beneficiarse del uso de los sistemas de organización de forma consciente, empleando algún tipo de interfaz que permita interactuar con ellos o de manera inconsciente, cuando algún sistema hace uso de los mismos de forma transparente para el usuario; esto es, el usuario desconoce que un determinado sistema con el que este está interactuando hace uso de los sistemas de organización del conocimiento para realizar una determinada tarea. Por ejemplo, un buscador que emplea un tesoro de forma automática para mostrar búsquedas relacionadas, de forma que el usuario no tiene la posibilidad de conocer el mecanismo por el que aparecen los resultados de búsqueda, ni interactuar de otra forma con el sistema de organización del conocimiento subyacente.

La investigación relativa al desarrollo de interfaces, que permiten al usuario emplear los sistemas de organización del conocimiento de forma consciente tiene un largo recorrido y múltiples estudios muestran las ventajas que este tipo de interfaces, tanto las visuales como aquellas basadas en texto, pueden tener para el usuario (Gaona-García, Martín-Moncunill, & Montenegro-Marin, 2017). Sin embargo, a nivel práctico y tras años de investigación, el uso consciente de los sistemas de organización del conocimiento no está extendido, contradiciendo los beneficios que resultan a nivel teórico y en experimentos controlados.

El presente trabajo indaga sobre posibles causas de este desuso y contribuir aportando metodologías que permitan evaluarlas, proporcionando información de utilidad para modificar un KOS de forma que se potencie su uso.

1.2 Planteamiento del problema

1.2.1 Motivación y oportunidad de investigación

Las aparentes ventajas que las interfaces basadas en KOS podrían tener para los usuarios fomentó que en el contexto del proyecto europeo de investigación Organic.Edunet (Palavitsinis, Ebner, Manouselis, & Sanchez-Alonso, 2011) se desarrollaran distintos tipos de interfaces que facilitaran este uso, atendiendo al tipo de tarea a la que se iba a dar soporte y al perfil y objetivos de los usuarios.

En concreto, el proyecto Organic.Edunet involucró a instituciones de 10 países para el desarrollo de un portal educativo en 18 idiomas, que proporcionara acceso a más de 10.000 recursos de aprendizaje sobre Agricultura Ecológica y Agroecología. Como parte del proyecto se elaboró una ontología (Sánchez-Alonso, y otros, 2008) y un perfil de metadatos (Palavitsinis, Manouselis, & Sánchez-Alonso, 2009) para la elaboración del sistema de información que manejaba los recursos que el repositorio contenía.

A continuación, se describen las interfaces empleadas en Organic.Edunet, indicando el tipo de KOS empleado. El portal actual, desarrollado en el marco del proyecto europeo de investigación “Organic.Lingua” ya no incluye estas interfaces, por lo que al final de la obra se anexan capturas de pantalla de las mismas.

1. **Búsqueda textual:** Empleaba las palabras claves que el usuario escribe, como parte del título y la descripción de todos los recursos educativos federados. Para facilitar acotar las búsquedas, existía una opción avanzada de la interfaz que permitía a los usuarios filtrar los resultados obtenidos de acuerdo al tipo de recurso educativo (vocabulario controlado) o su nivel educativo (vocabulario controlado) y ordenarlos según distintos criterios.
2. **Exploración:** Permitía explorar todos los recursos educativos federados de una forma polifacética. La exploración se producía de acuerdo a los tipos de recursos existentes (vocabulario controlado), los formatos de archivo (vocabulario controlado), los idiomas de los recursos (vocabulario controlado) y los niveles educativos (vocabulario controlado). Existía una interfaz textual y otra visual, ambas basadas en la taxonomía sobre la que se desarrolló la ontología de Organic.Edunet.
3. **Basada en etiquetas:** Facilitaba la localización de recursos de acuerdo a cómo otros usuarios los habían etiquetado (empleando los términos de la ontología de Organic.Edunet) o puntuado (en una escala del 1 al 5) previamente.
4. **Nube de etiquetas:** Existía una interfaz de nube de etiquetas en 2D en la portada de la web que daba acceso directo a las etiquetas que eran más solicitadas por los usuarios.
5. **Navegación semántica:** Búsqueda inteligente construida en base a la ontología de Organic.Edunet que permitía a los usuarios localizar recursos de acuerdo a los conceptos existentes en la ontología. Para apoyar a este método de búsqueda se desarrolló una interfaz gráfica que iba mostrando las relaciones ontológicas según el usuario clicaba sobre los términos.
6. **Otros:** El usuario podría llegar a otros resultados de búsqueda, a través de resultados relacionados. También existía una interfaz de búsqueda específica tipo textual para un conjunto de recursos que se consideró de gran relevancia, los llamados “Escenarios Educativos”.

El enfoque del consorcio que impulsó el proyecto se basó en que las distintas interfaces implementadas respondían a distintas necesidades de los usuarios y, en función de la tarea que estos tuvieran que desarrollar, elegirían la interfaz más adecuada para llevarla a cabo con éxito. Por otro lado, el desarrollo de las interfaces del portal estuvo basado en las recomendaciones que se podían extraer de la investigación sobre KOS, interfaces de búsqueda y visualización existentes en ese momento; con el fin de garantizar el mejor resultado posible para el proyecto.

Tras su implementación se realizó un estudio de usabilidad (Martín-Moncunill, 2012) en el que se evaluaban las distintas interfaces implementadas. Este estudio partió de una revisión heurística tras la que se realizó un profundo análisis centrado en la realización de tests con usuarios y el empleo de cuestionarios, entrevistas y grupos de enfoque como técnicas de apoyo. Las estadísticas de uso de las interfaces previas al estudio no mostraron gran aceptación de las mismas por parte de los usuarios, excluyendo la búsqueda textual.

Los resultados del mencionado estudio evidenciaron varios problemas de usabilidad en las interfaces que podrían ser resueltos de forma relativamente simple, pero apuntaron a una serie de problemas subyacentes, ligados al empleo que se estaba haciendo de los KOS, que suponían una gran barrera para su uso. A continuación, se presentan estos problemas, los cuales llevaron a las preguntas de investigación de la presente tesis doctoral. En el capítulo 4, sección 5 ("Selection and Use of Search Mechanisms in Learning Object Repositories: the Case of Organic.Edunet") se proporciona mayor detalle sobre este estudio e investigaciones posteriores relacionadas con el mismo.

1. A juicio de los usuarios existían recursos que no estaban catalogados adecuadamente. Los recursos procedían de distintas fuentes y su catalogación empleando la ontología creada para el proyecto y realizada de forma manual, no pareció resultar exitosa.
2. Muchos usuarios no entendían la ontología en sí misma, ni siquiera cuando se presentaba visualmente como una taxonomía, organizada sólo en base a las relaciones jerárquicas. La mayor parte de los comentarios recogidos durante los test de usabilidad y las entrevistas se centraron en:
 - a. Existían términos que no reconocían como propios del dominio.
 - b. La jerarquía empleada (de más genérico a más específico) no les parecía que hubiera sido desarrollada correctamente o directamente no la entendían. Varios usuarios consideraron que en algunas áreas faltaban niveles de jerarquía adicionales y que el nivel de especificidad entre los distintos niveles de la jerarquía resultaba poco uniforme.
3. La visualización de las interfaces podría ser mejorada, para facilitar no solo el uso de la misma y la comprensión de la ontología sino también permitir al usuario saber cuál era la utilidad de la interfaz. Gran parte de los usuarios manifestaron no percatarse de la utilidad (para qué servían) de cada una de las distintas interfaces y que sólo a través de su uso habían terminado por hacerse una idea del mismo.

4. En base a lo anterior se evidenció que los usuarios requerirían cierto tiempo de aprendizaje para hacer uso de las interfaces que les permitían interactuar directamente con el KOS; pero habría que tener en cuenta que, en un entorno real, lo más probable es que si un usuario no es capaz de reconocer la utilidad de una interfaz, directamente no invertiría tiempo en intentar o aprender a usarla.

1.2.2 Preguntas de investigación

Gracias al estudio de usabilidad realizado en Organic.Edunet, se observó la existencia de causas identificables por las que el uso y la utilidad de los KOS en el portal estaba limitado y surgió la idea de que los problemas detectados en el portal estuvieran relacionados con el hecho de que la adopción de interfaces que permitan hacer uso directo de KOS por parte de los usuarios no se haya extendido en la forma que podría estimarse en relación a los potenciales beneficios observados en los estudios existentes.

El objetivo principal de esta investigación es profundizar en las causas detectadas que puedan estar afectando a la adopción de los KOS en entornos reales, planteando metodologías que permitan evaluar la afectación de las mismas. Partiendo de este objetivo y considerando los resultados del estudio que dieron origen a esta tesis se plantean las siguientes preguntas de investigación:

1. ¿Las interfaces basadas en KOS tienen potencial de uso según los estudios existentes?

El punto de partida y la primera cuestión a analizar es plantearse si efectivamente existe, al menos a nivel teórico, potencial de uso en las interfaces basadas en KOS. Este potencial de uso se refiere a ventajas en cuanto a la utilidad que puedan aportar interfaces que empleen KOS y que, por tanto, motivarían al uso de las mismas. Para contestar a esta pregunta se ha trabajado en dos áreas:

- a) El análisis de la literatura sobre interfaces basadas en KOS, actualizado de forma permanente, la cual se recoge en el capítulo 2, que desarrolla el estado de la cuestión.
- b) La realización de varios experimentos sobre interfaces y visualización de KOS enfocados principalmente en el ámbito de la usabilidad. Estos experimentos se realizaron de forma paralela y continua a los que se desarrollaron para tratar de contestar al resto de preguntas de investigación, de manera que se pudo continuar verificando la influencia que la usabilidad del propio KOS tenía en las interfaces que eran ensayadas. Los experimentos realizados se resumen en el capítulo 4.

2. ¿Los algoritmos de catalogación automática de recursos constituyen una solución viable, proporcionando un nivel de calidad adecuado, frente al problema que puede constituir la catalogación manual?

En el anteriormente mencionado estudio de usabilidad del portal Organic.Edunet, la catalogación de recursos fue considerada en repetidas ocasiones como inadecuada. Esta catalogación se realizó de forma manual, empleando la ontología creada para el proyecto; la cual, por otro lado, también planteó notables problemas a los usuarios.

La catalogación manual de recursos presenta varias barreras (Park J. R., 2009) que dificultan garantizar la calidad de la misma, entre las que destacan:

- a) Requiere de expertos que conozcan el dominio y la ontología. Además, es de interés que los catalogadores tengan en cuenta el contexto de uso de su catalogación, de cara a optimizar la usabilidad de la misma. Cuando se trata de ontologías de reciente creación, como era el caso de Organic-Edunet, este asunto reviste todavía de mayor complejidad.
- b) Supone una gran cantidad de tiempo. Los catalogadores deben revisar los recursos y posteriormente valorar cuáles son los términos más adecuados del KOS para realizar la catalogación.
- c) Como consecuencia de esta suma de necesidad de experiencia y tiempo, aparece asociado un alto coste.

Teniendo en cuenta estos tres puntos, resulta evidente que la catalogación de recursos puede constituir un problema de entrada para el uso de KOS. Como alternativa a la catalogación manual de recursos han aparecido distintos algoritmos (Medelyan & Witten, 2005) que permiten la extracción de palabras clave haciendo uso de KOS, los cuales permitirían catalogar colecciones de forma mucho más rápida y ahorrando gran cantidad de esfuerzo tanto a nivel de coste como de esfuerzo.

La evaluación de la efectividad de este tipo de algoritmos se ha hecho típicamente utilizando medidas de precisión y exhaustividad (Lim, Wong, & Lim, 2013), que comparan la lista de palabras clave generadas a las listas proporcionadas por el autor. Esto supone únicamente una evaluación de la exactitud del sistema para hacer coincidir las palabras clave proporcionadas por el autor original, lo cual no termina de garantizar a los responsables de las colecciones de recursos la calidad y la aplicabilidad práctica de estas técnicas.

Para esto sería necesario un estudio más completo, que profundizara en la calidad relativa de las palabras clave asignadas por catalogadores humanos en comparación con las asignadas de forma automática y enfrentar directamente en términos de calidad el trabajo realizado por humanos con el de los algoritmos.

La respuesta a esta pregunta se estudia en la sección “Evaluación de la aplicabilidad práctica de la extracción automática de palabras clave”, dentro del capítulo 3.

3. ¿Evaluar la especificidad de dominio en los términos que componen un KOS podría tener impacto en su uso?

La motivación principal de la creación de un KOS suele residir en la necesidad de organizar la información y limitar la complejidad de un dominio concreto, de manera que el KOS pueda ser aplicado para resolver problemas dentro de este dominio. De esta forma, la construcción de los KOS suele partir de la identificación de una serie de términos (Cabré, 2005) propios del dominio (Hjørland B. , 2002) sobre los que se implementan distintas relaciones.

Estas terminologías establecidas para un propósito o dominio en particular se organizan normalmente mediante la utilización de algún tipo de clasificación (Hjørland, 2001) a partir de la cual se generan diferentes sistemas de organización del conocimiento (Hodge, 2000) específicos de un dominio.

Es frecuente que en este tipo de KOS de dominio específico, sobre todo aquellos de gran tamaño, se encuentren cantidades notables de términos generalistas o que puedan ser empleados en múltiples contextos, limitando la utilidad y la usabilidad del KOS. Por otro lado, estos términos específicos y genéricos podrían ser útiles por sí mismos (Kim y Cavedon, 2011), pero para ello sería necesario identificarlos.

Con el fin de poder evaluar la especificidad de dominio de los KOS y sus términos se propone un estudio empleando el tesoro AGROVOC, que se recoge en la sección “Evaluación del grado de especificidad de dominio en los términos de un KOS”, dentro del capítulo 3.

4. ¿Asignar un grado de distancia de especialización entre conceptos podría mejorar la usabilidad de un KOS?

La relación de generalización / especialización ("gen-spec") es una relación entre clasificadores (términos o conceptos) que implica una relación taxonómica y su subsiguiente semántica hereditaria. Esta relación jerárquica es empleada en los tesauros con la denominación inglesa

“broader / narrower” (ISO 25964, 2013). Las ontologías usan la relación conocida como “is-a” para representar el mismo tipo de relación.

El entendimiento común de este tipo de relaciones jerárquicas las considera como “todo o nada”, asumiendo que la relación siempre es igualmente fuerte entre un clasificador y cualquiera de sus clasificadores relacionados mediante una relación “gen-spec” y también en todos los niveles de la jerarquía, lo cual supone para algunos autores una simplificación excesiva del relato psicológico de las relaciones del mundo real (Sicilia, García, Aedo, & Díaz, 2003).

Este hecho ha sido considerado como problema por autores como Cohen y Murphy (2003), Sicilia y otros (2003), Cross (2004) o Hu y otros (2007), ante el cual se han propuesto soluciones basadas en la asignación de un grado de distancia relativa para representar el nivel de similitud entre los pares de clasificadores relacionados (Sicilia, García, Aedo, & Díaz, 2003).

Este tipo de jerarquía fue la que se empleó para representar la ontología de Organic.Edunet, respecto a la cual los usuarios señalaron que a su juicio existían inconsistencias, como la falta de niveles de jerarquía adicionales o que el nivel de especificidad entre los distintos niveles de la jerarquía resultaba poco uniforme.

En base a lo expresado por estos autores, asignar una distancia de especialización entre conceptos podría resultar de utilidad para mejorar la usabilidad de un KOS. En la sección “Evaluación de la distancia de especialización entre conceptos”, dentro del capítulo 3, se propone una metodología para asignar esta distancia de especialización a un área del tesauro AGROVOC, para posteriormente indagar sobre su utilidad en dos prototipos de interfaces, una textual y otra visual.

1.2.3 Objetivos

El objetivo principal de este trabajo es indagar sobre algunas de las posibles causas que están afectando a la adopción de los KOS, en relación al potencial observado a nivel teórico y contribuir aportando metodologías que permitan evaluar estas causas, proporcionando información de utilidad para modificar un KOS de forma que se potencie su uso desde el enfoque del usuario.

En concreto, se profundizará en la evaluación y análisis de las siguientes posibles causas, cuya identificación y preguntas de investigación relacionadas se han introducido en las secciones anteriores; a partir de las cuales se originan los siguientes objetivos específicos:

- a) La catalogación automática de recursos empleando KOS de dominio específico.

- i) Garantizar a los responsables de las colecciones de recursos una solución, para evaluar la calidad de las palabras clave extraídas automáticamente, más adecuada al tradicional criterio de precisión/exhaustividad.
 - ii) Emplear este mecanismo de evaluación en un caso real y aportar los resultados para comprobar si la catalogación automática de recursos para el algoritmo analizado sería una solución viable.

- b) La evaluación de la especificidad de dominio en los términos que componen los KOS.
 - i) Proporcionar un mecanismo de evaluación de la especificidad de dominio de un KOS.
 - ii) Aplicar este mecanismo para obtener indicadores que permitan:
 - (1) establecer el grado de especificidad de dominio global de un KOS.
 - (2) clasificar los términos del KOS en función a su grado de especificidad de dominio.
 - (3) analizar el esfuerzo empleado en la aplicación de la metodología.
 - iii) Indagar sobre la posibilidad de emplear mecanismos automatizados para facilitar la evaluación de especificidad de dominio de un KOS y categorizar sus términos.

- c) Las posibilidades de uso de la distancia de especialización entre conceptos, tanto en interfaces visuales como basadas en texto.
 - i) Proponer un sistema que permita asignar valores de distancia de especialización entre conceptos a KOS.
 - ii) Aplicar este sistema para obtener un KOS que incluya información sobre la distancia de especialización entre sus conceptos.
 - iii) Emplear el KOS obtenido para realizar una primera evaluación sobre las posibilidades de uso de la distancia de especialización entre conceptos en:
 - (1) Interfaces basadas en texto.
 - (2) Interfaces visuales.

1.4 Metodología de trabajo

El presente trabajo incluye la propuesta de metodologías para la evaluación y el análisis de varios aspectos ligados a la usabilidad de los sistemas de organización del conocimiento, centrándose en aquellos contruidos sobre un dominio específico y su uso directo mediante interfaces. Para ello se parte fundamentalmente de las metodologías y técnicas existentes en las áreas de usabilidad (Nielsen J. , 1994) y la interacción persona-ordenador (Jonathan, Jinjuan, & Harry, 2009), incluyendo tanto las orientadas al análisis cualitativo como al cuantitativo (Rubin & Dana, 2008).

La representación de KOS mediante interfaces aparece como una herramienta que permite a los usuarios emplear los KOS de manera más amigable, constituyendo una de las vías de

potenciación del uso de los mismos. A este respecto se han realizado varias propuestas que, como se indicó anteriormente, parecen haber tenido una aceptación muy limitada. Es por ello que la evaluación de las interfaces que permiten la visualización de los esquemas generados a partir de los términos y relaciones de los KOS ocupa gran parte del análisis de la literatura desarrollado para el presente trabajo, que se contrastará con los resultados de experimentos propios desarrollados en el contexto de diversos proyectos europeos de investigación, resumidos en el capítulo 4.

Uno de los motivos que pudiera dificultar la adopción del uso de los KOS en tareas relativas a la localización de información, es la necesidad de relacionar los recursos existentes en cualquier tipo de repositorio de información, con los términos pertenecientes al KOS. Realizar esta catalogación de forma correcta es el primer paso para poder aprovechar el potencial de los KOS, para lo cual se emplean KOS de dominio específico, que permitan clasificar con el detalle suficiente los distintos recursos. Sin embargo, cuando esta tarea se desarrolla mediante expertos humanos resulta muy costosa, lo que ha incentivado el desarrollo de diferentes mecanismos para que sea realizada de forma automática.

Atendiendo a esto, se realizará un estudio que comprenderá la investigación sobre algoritmos que permiten la extracción de palabras clave a partir de textos para su catalogación. Este estudio está directamente relacionado con el objetivo “a”, esto es, la catalogación automática de recursos empleando KOS de dominio específico.

Aunque existen KOS que no se centran en un dominio de conocimiento o discusión específico como SUMO (Niles & Pease, 2001) o MILO (Niles & Terry, 2004) la creación de KOS suele venir motivada por la necesidad de organizar la información y limitar la complejidad de un dominio concreto, de manera que el KOS pueda ser aplicado para resolver problemas dentro de ese dominio. Sin embargo, los KOS específicos de dominio pueden contener términos generalistas o que puedan ser empleados en múltiples contextos, de manera que a partir de estos términos no sería posible identificar el dominio del KOS, limitando la utilidad y la usabilidad del KOS.

Con el fin de profundizar en esta problemática se planteará un estudio que analizará la especificidad de dominio de los términos de los KOS, proponiendo un método para su evaluación. Este estudio está directamente relacionado con el objetivo “b”, esto es, la evaluación de la especificidad de dominio en los términos que componen los KOS.

Las relaciones del tipo “Generalización/Especialización” son relaciones jerárquicas usadas habitualmente para organizar los conceptos existentes en los KOS, como en la relación “broader-narrower”, que encontramos en las taxonomías o la relación “is-a” que sería su equivalente en las ontologías. El entendimiento común de este tipo de relaciones asume que la relación siempre es igualmente fuerte entre un concepto y cualquiera de sus conceptos relacionados y también en todos los niveles de la jerarquía. Esto supone una simplificación

excesiva de las relaciones del mundo real, la cual podría ser limitada asignando valores a las relaciones que identifiquen la distancia de especialización entre los conceptos.

Siguiendo este razonamiento, se llevará a cabo un estudio para analizar los beneficios del empleo de esta “distancia de especialización de concepto” mediante una evaluación con usuarios, aplicando la propuesta en una interfaz textual y otra visual. Este estudio está directamente relacionado con el objetivo “c”, esto es, las posibilidades de uso de la distancia de especialización entre conceptos, tanto en interfaces visuales como basadas en texto.

1.5 Estructura del documento

Esta memoria de tesis doctoral se encuentra dividida en los siguientes capítulos:

1. El primer capítulo introduce el tema a tratar y la motivación, indicando el objeto de estudio, el planteamiento del problema, las preguntas de investigación, los objetivos, la metodología de trabajo y, finalmente, anexa los artículos publicados, participación en proyectos de investigación y estancias internacionales relacionadas con la tesis.
2. En el segundo capítulo se presenta el estado de la cuestión, donde se exponen de manera detallada cada uno de los temas tratados en la tesis, a saber: características y tipos de KOS, especificidad de dominio, distancia de especialización entre conceptos, técnicas de extracción automática de palabras clave, interfaces y visualización.
3. En el capítulo tercero se desarrollan y aplican los mecanismos de evaluación y análisis de los factores limitantes observados.
4. En el capítulo cuarto se resume el trabajo realizado en los estudios relativos a interfaces visuales que emplean KOS, cuyos resultados respaldan la existencia de los problemas de usabilidad de KOS tratados en el capítulo anterior.
5. Finalmente, en el capítulo quinto se exponen las conclusiones, relacionándolas con los objetivos, las preguntas de investigaciones y las aportaciones realizadas.

Capítulo 2: Estado de la cuestión

“I must not fear. Fear is the mind-killer. Fear is the little-death that brings total obliteration. I will face my fear. I will permit it to pass over me and through me. And when it has gone past I will turn the inner eye to see its path. Where the fear has gone there will be nothing. Only I will remain.”

(“Letanía contra el miedo” de la obra “Dune”, de Frank Herbert, 1965)

2.1 Introducción sobre Sistemas de Organización del Conocimiento

El término “Knowledge Organization Systems” (KOS) o “Sistemas de Organización del Conocimiento” fue definido por el grupo de trabajo “Networked Knowledge Organization Systems Working Group”, en su primera reunión en la conferencia de la ACM “Digital Libraries '98 Conference” en Pittsburgh, Pennsylvania: *“Los sistemas de organización del conocimiento (KOS – “Knowledge Organization Systems”) incluyen todos los tipos de esquemas para la organización del conocimiento”.*

Como ya se comentó en la introducción, existen iniciativas para construir sistemas de organización del conocimiento generalistas, pero los KOS están mayoritariamente concebidos para ayudar a crear, gestionar y visualizar modelos que muestren una perspectiva simplificada de los conceptos que forman parte de un determinado dominio y de la estructura semántica subyacente (Nafría, Alemany, & Pérez-Montoro, 2017).

Los KOS parten de una base de términos sobre los cuales se establecen distintas relaciones, que dotan a la construcción de una cierta carga semántica (Daconta, Obrst, & Smith, 2003), dando lugar a distintos tipos de sistemas como los tesauros, las ontologías, los mapas conceptuales o las listas de sinónimos. Cada uno de estos sistemas presentará una complejidad distinta, existiendo estructuras como los tesauros que están normalizadas, en el sentido que cuentan con una estructura marcada y un tipo de relaciones determinadas.

Así, una de las formas más elementales de uso de los KOS se basa en emplear los términos como palabras clave que faciliten la clasificación de la información y permitan acotar resultados. De esta forma, los KOS como herramienta para la clasificación, organización y localización de información, se han demostrado útiles ya desde antes de la aparición de los sistemas informáticos.

Las relaciones que se establecen entre los términos de los KOS y la carga semántica que se construye permite realizar razonamientos útiles para la creación de funcionalidades avanzadas que resultan de gran interés para áreas como la inteligencia artificial y otras técnicas de análisis de datos, que aprovechan la información subyacente en los distintos KOS (Tickoo & Iyer, 2017).

Estas relaciones también permiten que los KOS sean plasmados de forma visual, de manera que los usuarios pueden beneficiarse de esta información de manera consciente, empleando interfaces que faciliten tanto la comprensión del KOS como su uso. Además, el hecho de navegar por un KOS puede ser en sí mismo una herramienta de aprendizaje, dado que mediante el estudio de los términos incluidos y sus relaciones se puede extraer conocimiento sobre el dominio tratado.

Como se analiza en las siguientes secciones de este capítulo, la investigación en cuanto al desarrollo de interfaces (visuales y no visuales) que permitan al usuario final operar con KOS es extensa y las conclusiones de los estudios muestran múltiples ventajas y posibilidades de empleo de las mismas.

Sin embargo, hasta la fecha, este tipo de interfaces no se han extendido y apenas son ofrecidas y empleadas para el uso que a nivel teórico debiera existir. Por el contrario, es frecuente que las herramientas de búsqueda actuales hagan uso de esta información, pero sin indicarlo de forma explícita y sin permitir al usuario que profundice en el uso de los KOS, limitando sus posibilidades.

El objeto principal de esta obra es precisamente investigar en relación a causas identificables por las que el uso y la utilidad de los KOS está limitado en este ámbito (uso consciente) y no se extiende en la forma en la que se ha venido reflejando en multitud de artículos de investigación.

Las siguientes secciones de este capítulo analizan el estado del arte sobre los que se desarrollan las preguntas de investigación objeto de la presente obra, a saber:

- 1) Teoría de dominio y KOS de dominio específico.
- 2) Extracción automática de términos o palabras clave.
- 3) Distancia de especialización entre términos.
- 4) Visualización e interfaces visuales de KOS, profundizando en su uso en repositorios y librerías digitales.

2.2 Principales tipos y usos de KOS

Los sistemas de organización del conocimiento se han mostrado de gran utilidad para diferentes tareas relacionadas con la recuperación de la información y los propósitos de búsqueda (Lancaster, 1972), tales como anotar e indexar los recursos utilizando vocabularios controlados (Park, y otros, 2013), expandir las consultas de búsqueda realizadas por usuarios empleando técnicas de expansión de consultas basadas en el uso de ontologías (Segura, García-Barriocanal, & Prieto, 2011), posibilitar la realización de búsquedas inteligentes empleando semántica (Madhu, Govardhan, & Rajinikanth, 2011) (Sicilia, 2014) o deducir el tema, en el sentido descrito por (Hjørland B. , 2001), de un documento (Paranjpe, 2009) (Gamon, Yano, X, Apacible, & Pantel, 2013).

De hecho, la información contenida en un KOS es fácilmente interpretable por máquinas y por tanto de utilidad para el campo de la inteligencia artificial. Desde los comienzos de este campo, los investigadores centrados en el mismo han reconocido que la captura de conocimientos es una pieza clave para la construcción de sistemas de inteligencia artificial y se percataron del potencial de las ontologías para crear modelos computacionales que permitan ciertos tipos de razonamiento automatizado (Chandrasekaran, Josephson, & Benjamins, 1999).

Si bien la información contenida en los KOS puede ser empleada automáticamente por sistemas, sin que se requiera supervisión ni acción alguna por parte del usuario; sí existen interfaces que permiten interactuar y emplear los KOS de manera consciente. Entre estas, los principales esfuerzos se han dirigido al desarrollo de interfaces visuales que permiten el uso de estos sistemas para formular consultas (García-Barriocanal & Sicilia-Urbán, 2003) mediante la selección de términos y especificación de relaciones, siguiendo las técnicas descritas por MJ Bates (1989), con la posibilidad de realizarlo navegando a través de los mismos (Gaona García, Martín-Moncunill, Sánchez-Alonso, & García, 2014).

Los distintos tipos de KOS emplean distintas notaciones, que determinan aspectos relacionados con el área temática, niveles de relación, vínculos y formas de asociación, entre otras. Atendiendo a estas características, variará el uso que se pueda hacer del KOS. A continuación se mencionan los KOS más relevantes, para un mayor detalle se recomienda consultar “glosariumBITri” (Nafría, Alemany, & Pérez-Montoro, 2017); dónde los autores proporcionan y mantienen actualizada información exhaustiva al respecto.

Diccionarios y Glosarios

Estos KOS recogen términos que se encuentran ordenados alfabéticamente proporcionando su definición y significado junto con información adicional como etimología, ortografía, pronunciación, separación silábica y forma gramatical. Los glosarios recopilan definiciones de términos contextualizados en un tema o disciplina, mientras que el diccionario no presenta esta acotación.

Vocabularios controlados

Constituyen un conjunto de términos que tienen asignado un significado concreto (Cumming, 2003). La norma ANSI/NISO Z39.19 2005 (Rev. 2010) (ANSI/NISO, 2005) recoge cuatro tipos de clasificación sobre vocabularios controlados: listas, anillos de sinónimos, taxonomías y tesauros.

1. **Lista:** Set de palabras o frases dispuestos en series organizadas.
2. **Anillo de sinónimos:** Set de palabras o frases que son consideradas como equivalentes para los propósitos relacionados con la recuperación de información.
3. **Taxonomía:** Lista estructurada o árbol de palabras o frases, organizada de manera jerárquica desde los términos más generales a los más específicos, incluyendo términos relacionados. La palabra procede del griego (τάξις) taxis = ordenación y (νομος) nomia = reglas, reglas de ordenación o clasificación, siendo un concepto originalmente asociado a la rama de la biología, usado para la clasificación de organismos, plantas, géneros, reinos, formas, etc., donde varias especies se encuentran agrupadas de acuerdo a un conjunto de similitudes percibidas a partir de su forma, orden o clase.
4. **Tesauro:** Un conjunto de palabras o frases con términos equivalentes identificados explícitamente, de manera que se elimine la ambigüedad y garantice que cada concepto se contextualice de manera inequívoca. Este tipo de KOS incluye relaciones de equivalencia, jerarquía y asociativa; muestran conexiones laterales: términos relacionados, términos no preferentes y preferentes y son polijerárquicos, conteniendo notas de alcance para indicar el significado de algunos términos.

Ontologías

La construcción de ontologías es uno de los acercamientos más empleados para plasmar formalmente el conocimiento de un dominio. En los 90 se consolidan varias iniciativas que se centran en la construcción de ontologías desde el enfoque de la ingeniería, y que acuñan el término de "*Ontology Engineering*". Por mencionar un par de ejemplos, en 1997 los investigadores españoles Fernández-López, Gómez-Pérez, & Juristo, (Fernández-López, Gómez-Pérez, & Juristo, 1997) proponen una de las metodologías de mayor relevancia a este respecto. Poco después los investigadores japoneses, relacionados con el campo de la inteligencia artificial, Mizoguchi e Ikeda (1998), inciden en la creación de reconocer un nuevo campo de investigación denominado "*Ontology Engineering*".

En la actualidad esta idea parece implantada, situación que se puede ver reflejada en la creación de distintas metodologías relativas al desarrollo de ontologías que se refieren específicamente a “*Ontology Engineering*”; como HCOME (Human-centered ontology engineering) (Kotis & Vouros, 2006), IOEM (Ontology engineering methodology for large systems) (Sliwa, Gleba, Chmiel, Szwed, & Glowacz, 2011), DILIGENT (DIstributed, Loosely-controlled and evolvinG Engineering of oNTologies) (Pinto, Staab, & Tempich, 2004), DOGMA (Jarrar & Meersman, 2009), o la reciente propuesta de De Nicola y Missikoff (2016) centrada en el desarrollo ágil de ontologías.

La definición del término ontología ha evolucionado durante las últimas décadas (Gómez-Pérez, Fernández-López, & Corcho, 2004); entre las definiciones más conocidas se encuentran la realizadas por Neches y otros, (1991) “una ontología define los términos básicos y relaciones que conforman el vocabulario de un área específica, así como las reglas para combinar dichos términos y las relaciones para definir extensiones de vocabularios”; y la de Gruber (1993), que define una ontología como “una especificación explícita de una conceptualización”.

Adicionalmente (Gruber, 1993) propone los siguientes principios para mejorar el diseño de ontologías en cuanto a la comunicación del conocimiento:

- Claridad: definiciones neutras, formalizadas con axiomas y completas, con condiciones necesarias y suficientes.
- Mínima codificación
- Extensibilidad
- Mínimo compromiso ontológico: menor número de suposiciones posible.

Estos principios son extendidos por (Gómez-Pérez, Fernández-López, & Corcho, 2004), quien propone:

- Declarar los elementos disjuntos y expresar el conocimiento de forma exhaustiva
- Minimizar la distancia semántica de los elementos que están al mismo nivel de jerarquía
- Estandarizar las denominaciones de los elementos de forma comprensible

Según (Nafría, Alemany, & Pérez-Montoro, 2017) uno de los motivos que subyacen bajo la poca convergencia de las definiciones es el de encontrar una definición que agrupe a todos los recursos propuestos hasta el momento como ontologías. Por ejemplo, desde la perspectiva de la inteligencia artificial las ontologías se contemplan como un instrumento que define una serie de términos básicos a partir de un vocabulario en un área de conocimiento, junto con una serie de reglas de vinculación entre términos y relaciones para definir extensiones a este vocabulario (Corcho, Fernández-López, & Gómez-Pérez, 2003).

Las ontologías pueden clasificarse de acuerdo a su forma, estructura y uso (Valente, 2005), entre ellas cabe mencionar las siguientes por su relación con el presente trabajo; posteriormente se dará mayor detalle sobre las mismas:

- Las ontologías para representación de conocimiento (Davis, Shrobe, & Szolovits, 1993) (Gruber, 1993) (Guarino & Poli, 1995) permiten relacionar conceptualizaciones asociadas a formalismos de representación de conocimiento.
- Las ontologías de dominio contienen todos los conceptos asociados a un dominio (Hjørland & Albrechtsen, 1995) particular, a partir de una estructura detallada de un contexto y subdominios asociados (Mizoguchi, Vanwelkenhuysen, & Ikeda, 1995).
- Las ontologías de alto nivel (*en inglés “upper-level ontologies” o “top ontologies”*), también se hace referencia a ellas como “ontologías fundacionales” y “ontologías de propósito general”. Como su propio nombre indica, son ontologías que describen conceptos muy generales y son comúnmente empleadas para interrelacionar ontologías entre sí empleando sus elementos raíz (Mascardi, Locoro, & Rosso, 2009).

Gaona García (2014) recoge en su tesis doctoral una imagen que presenta a modo resumen, una adaptación de la propuesta realizada por Lei (2008), sobre el soporte que presentan algunos esquemas de representación de conocimiento en cuanto a su estructura y organización de contenidos. Esta imagen se reproduce a continuación (figura 1) por la claridad con la que refleja el tipo de información que contienen los distintos KOS, lo cual define el tipo de uso que se les puede dar.

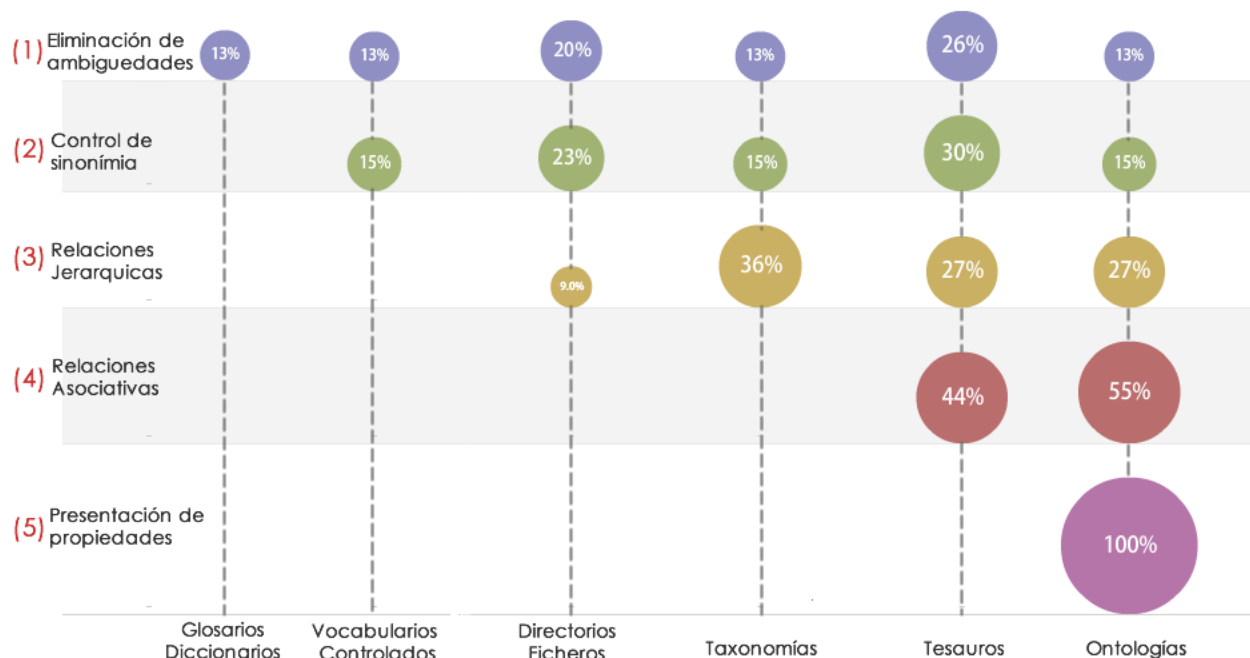


Figura 1. Estructuras y estrategias para organización del conocimiento. Extraída de (Gaona-García P. A., Visualización de esquemas de representación de conocimiento para el acceso a recursos en repositorios digitales, 2014) y basada en (Lei, 2008)

Independientemente del tipo de KOS, para que este sea efectivo debe ser consistente y lo más realista para representar un área temática y sus relaciones (Chrysafiady & Virvou, 2013), adecuándose al uso que se le va a dar.

Por otro lado, del mismo modo que un sistema dotado de inteligencia artificial puede realizar ciertos tipos de razonamiento empleando KOS, un ser humano sería capaz de extraer y asimilar conocimientos mediante el estudio de este. Es por ello que los KOS pueden considerarse en sí mismos fuentes valiosas de información con potencial para ser empleados como herramientas de aprendizaje.

En el contexto educativo, es un hecho ampliamente demostrado que el uso de técnicas de visualización tales como los diagramas son herramientas muy valiosas para fines de aprendizaje (Martín-Moncunill, Sánchez-Alonso, Gaona-García, & Marianos, 2013), del mismo modo, la visualización y navegación por un KOS podría no ser sólo útil para la localización y la indexación de recursos (Gaona-García, Sánchez-Alonso, & Montenegro, 2014), sino también como una experiencia de aprendizaje en sí (Gaona-García, Stoitsis, Sanchez-Alonso, & Biniari, 2016).

La figura 2 refleja razonamientos que pueden surgir al navegar por un tesoro sobre mamíferos, de forma similar a la que se detectó en experimentos sobre navegación por KOS realizados en el marco de la presente tesis doctoral y resumidos en el capítulo 4.

En esta figura los alumnos observan un KOS de mamíferos y según avanzan en su navegación, surgen distintas situaciones ligadas al contexto educativo, como pueden ser las mostradas en la figura 2, a saber:

- La adquisición de conocimiento en la alumna que no sabe que las marmotas son roedores.
- La reflexión al surgir un dilema en el alumno que tiene un conocimiento equivocado sobre la clasificación de los canguros
- La motivación para indagar más sobre un tema en la alumna que quiere saber más sobre los capibaras.

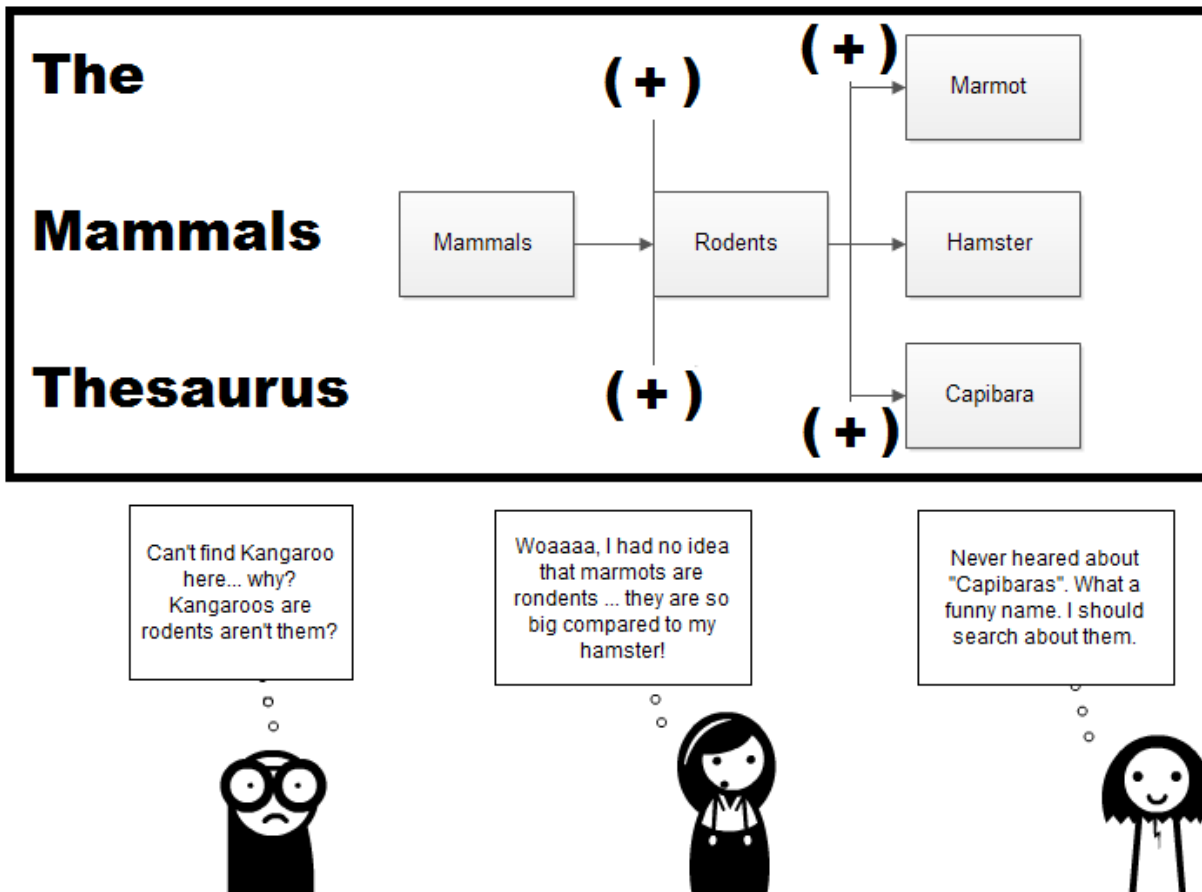


Figura 2. Razonamientos que pueden surgir al navegar por un tesauro en un contexto educativo en base a los experimentos realizados.

Es por ello que el empleo de KOS en repositorios digitales y especialmente en aquellos centrados en la educación, serían un caso paradigmático de la utilidad del empleo de interfaces que permitan al usuario hacer uso directo y consciente del KOS. En la sección de este capítulo dedicada a la visualización de KOS y su integración en repositorios digitales se profundizará en este aspecto.

2.2 KOS de dominio específico

Los KOS específicos de dominio contienen listas de términos específicos de dentro y fuera del dominio que podrían ser útiles por sí mismos (Kim & Cavedon, 2011). El primer paso para desarrollar un KOS específico de dominio, independientemente del método, es generar una colección de términos específicos de dominio. Kim y Cavedon (2011) definieron "términos específicos de dominio" como términos que tienen uno o más significados significativos en un dominio específico. Por tanto, estos KOS se basan en terminologías específicas de dominio y representan conceptos (Stock, 2010) que pertenecen a un dominio particular y cuyos términos tienen significados particulares aplicados a ellos en ese dominio.

Broughton (2006) considera que el análisis de facetas, del inglés: "facet analysis", (Classification Research Group, 1955) constituye el núcleo de la construcción de KOS, pero Mai (2006) argumenta que la atención debe centrarse en el análisis de dominios como la actividad que precede al análisis de facetas, recomendando la aplicación de técnicas de análisis cognitivo para la realización de esta tarea. El concepto de "análisis de dominio" fue introducido por Hjørland y Albrechtsen (1995), los cuales sostienen que el paradigma del análisis de dominio es entender las actividades y productos de un dominio particular.

La interpretación del enfoque analítico del dominio en el contexto de la ciencia de la información sugerida por Hjørland y Albrechtsen (1995) y Hjørland (2002) es seguida por la de Mai (2004), que afirma que, para construir cualquier tipo de esquema de indexación para una comunidad en particular, se debe partir del estudio de cómo se hacen las preguntas en ese ámbito. En artículos posteriores Mai (2005) propone un enfoque centrado en el dominio de la indexación como una alternativa al enfoque centrado en el documento, que sostenía bloquearía factores dependientes del contexto en el proceso de indexación. Sin embargo, Hjørland y Albrechtsen no definieron claramente el concepto de "dominio" y, según Mai (2005), podría no ser posible, ni siquiera deseable, definirlo de manera exclusiva. Mai (2005) considera que "el dominio es un concepto evolutivo y abierto que se desarrollará a medida que el concepto se utilice y se aplique en la investigación y la práctica".

El concepto de "dominio" ha sido ampliamente utilizado en las ciencias de la información y campos relacionados, pero no siempre con el mismo significado. Hjørland y Kyllesbech Nielsen (2001) incluyen "dominio" en una lista de términos que a veces se consideran sinónimos y a veces se considera que tienen diferentes significados. Más precisamente, consideran los términos ingleses "subject", "aboutness", "topic", "theme", "domain", "field", "content" e "information"; como los términos ambiguos más relevantes.

En su trabajo posterior, Hjørland (2001) proporciona un análisis más profundo sobre el uso y la discusión del significado y la naturaleza de estos términos. En cuanto al dominio, Hjørland considera que es un concepto moderno relacionado con la ciencia cognitiva y las tecnologías de la información. Continuando con su investigación sobre "análisis de dominio", Hjørland (2002) propone once enfoques para el análisis de dominios; estos enfoques incluyen el uso de clasificaciones especiales y tesauros centrados en los enfoques basados en facetas, estudios empíricos de usuarios y estudios terminológicos.

La utilidad de terminologías específicas de dominio ha aumentado el interés en el desarrollo de mecanismos automáticos para identificarlos y acumularlos. Se han descrito varios métodos de extracción para términos específicos de dominio; la mayoría de ellos aplican medidas estadísticas basadas en el uso para determinar la especificidad del dominio de los términos candidatos (Kim & Cavedon, 2011). Milne y otros (2006) utilizaron entradas de Wikipedia como términos específicos de dominio y revisaron los términos en AGROVOC (Sini, Lauser, Salokhe, Keizer, & Katz, 2008) con entradas de Wikipedia para verificar la especificidad de dominio de las entradas de Wikipedia. Kida y otros (2007) propusieron que un dominio puede ser identificado a través de una lista de términos de dominio conocidos, por ejemplo los clasificados por expertos humanos en AGROVOC. Vivaldi y Rodríguez (2010) consideraron que las categorías de Wikipedia eran específicas del dominio y las utilizaron para extraer los términos específicos del dominio bajo el dominio de destino.

Sin embargo, es común que los KOS específicos de dominio de gran tamaño, como AGROVOC, el cual abarca todos los ámbitos de interés de la Organización de las Naciones Unidas para la Alimentación y la Agricultura; incluyan términos que podrían ser relevantes o de uso común en el dominio, pero no son completamente específicos para ese dominio en el sentido de que no permiten identificar adecuadamente el verdadero dominio KOS. Por ejemplo, AGROVOC incluye términos tales como "contabilidad", "auditoría", "contabilidad de costos" y "depreciación", que podrían ser relevantes para aspectos del dominio agrícola, así como para otros dominios, a pesar de que caracterizan más estrechamente al dominio de la economía.

Este cruce entre términos y dominios puede crear algunos problemas cuando se trabaja con mecanismos automáticos que usan este tipo de KOS y asumen que todos sus términos son completamente específicos del dominio. Por ejemplo, teniendo en cuenta el número de términos relacionados con la economía en el tesoro AGROVOC, un texto puramente económico podría ser categorizado como un texto agrícola por un mecanismo automático que utilizara AGROVOC como patrón. Hasta donde sabemos, no existen enfoques previos que traten de identificar términos genéricos y específicos en terminologías extensas, específicas de dominio u otros KOS, que usualmente se tomen como estándares.

Optimizar y facilitar la recuperación de información, ya sea desde el punto de vista computacional o humano, mediante el uso de KOS es un tema de interés que ha sido abordado desde numerosas perspectivas. Estudios previos muestran cómo las relaciones y propiedades establecidas por KOS pueden apoyar la búsqueda inteligente mediante el uso de la semántica formal (Madhu, Govardhan, & Rajinikanth, 2011) y aumentar las consultas de los usuarios utilizando técnicas de expansión de consultas (Segura, García-Barriocanal, & Prieto, 2011).

La evaluación del grado de especificidad de dominio en terminologías de gran tamaño podría ser una herramienta útil para estos casos, permitiendo el refinamiento de técnicas de expansión de consultas utilizadas para mejorar la recuperación de información para fines de búsqueda y también optimizar las técnicas utilizadas para reducir la complejidad de clasificación de sujetos (Julien, Tirilly, Dinneen, & Guastavino, 2013), y así facilitar el proceso de navegación por KOS.

Por otro lado, las ontologías superiores son ontologías que describen conceptos muy generales que son los mismos en todos los dominios. Estas han sido explotadas por agentes de software inteligentes para tareas tales como fomentar la interoperabilidad entre los esquemas conceptuales de SKOS (Sanchez-Alonso & Garcia-Barriocanal, 2006), la adaptación de ontologías (Mascardi, Locoro, & Rosso, Automatic ontology matching via upper ontologies: A systematic evaluation., 2009) o el desarrollo de sistemas omnipresentes (Stevenson, Knox, Dobson, & Nixon, 2009).

Según Mascardi y otros, (2007), "las ontologías superiores pueden ser explotadas como una lingua franca por agentes de software inteligentes en todos aquellos escenarios donde es imposible (o no hay voluntad) que un agente revele su propia ontología a otro agente, a pesar de la necesidad de comunicarse con ella".

Actualmente, SUMO es la ontología de propósito general más extendida. Esta es propiedad del IEEE aunque de libre uso y consiste en varios módulos: el mismo SUMO, la Ontología de Nivel Medio (MILO) y diferentes ontologías de dominios específicos. SUMO se encuentra a su vez mapeada con WordNet (Niles & Pease, 2003), una gran base de datos léxica de inglés donde los sustantivos, verbos, adjetivos y adverbios se agrupan en conjuntos de sinónimos cognitivos, cada uno de los cuales expresa un concepto distinto y se interconecta a través de relaciones conceptuales, semánticas y léxicas.

2.3 Distancia de especialización entre conceptos

Como ya se ha indicado, la construcción de KOS tales como tesauros u ontologías, es una tarea compleja que normalmente comienza generando un corpus de términos (Kim & Cavedon, 2011). Tras esto, los términos se pueden estructurar estableciendo relaciones entre ellos. En este sentido, un KOS es una forma de representación del conocimiento que tiene como objetivo organizar la terminología para un propósito particular.

La relación de generalización / especialización ("gen-spec") es una relación entre clasificadores (términos) que implica una relación taxonómica y su subsiguiente semántica hereditaria. Según (Aitchison, Gilchrist, & Bawden, 2000) los tesauros emplean este tipo de relaciones jerárquicas ("broader/narrower" en inglés) proporcionando información adicional acerca de qué términos están relacionados (en inglés "related terms", representan relaciones asociativas) y qué términos se pueden utilizar como sinónimos (relaciones de equivalencia). De manera similar, las ontologías usan la relación conocida como "is-a" para representar jerarquías, comparable a la relación mencionada en los tesauros.

Una relación "gen-spec" existe entre dos entidades (también denominadas clasificadores, clases, sujetos, etc.) si una de las entidades evoca una especificidad o un caso más específico de la otra; por ejemplo, un hámster es un caso específico de roedor. La relación "gen-spec" es uno de los conceptos esenciales en el área de representación del conocimiento (Fotzo & Gallinari, 2004) y no sólo es ampliamente utilizado para la construcción de KOS, sino también en campos tales como la lógica, las notas de modelado orientadas a objetos de propósito general (Object Management Group (OMG), 2013) o los lenguajes de programación (Norrie, Reimer, Lippuner, Rys, & Schek, 1994).

La relación "gen-spec" permite construir una organización jerárquica de los conceptos que están presentes en un corpus de términos. Esto es útil no sólo para proporcionar una organización jerárquica de una colección de documentos sino también para facilitar tareas más complejas como los casos de predicciones (Cerri, Barros, Freitas, & de Carvalho, 2014) en la ontología "gene-ontology" (Khan, Baig, & Shahzad, 2014).

Fotzo y Gallinari (2004) afirmaron que, aunque otros tipos de relaciones pueden conducir a una organización jerárquica de conceptos y documentos, la relación "gen-spec" es muy intuitiva para los usuarios y es la relación más utilizada para estructurar colecciones.

De hecho, la utilidad de proporcionar y refinar las jerarquías de las colecciones ha sido un tema de investigación por sí mismo (Ryutaro, Hideaki, & Shinichi, 2001) (Morita, Tamagawa, & Yamaguchi, 2014), y varios autores han propuesto mecanismos semiautomáticos (Chodorow, 1985), así como enfoques automáticos (Dupret & Piwowarski, 2006) para realizar esta tarea.

Sicilia y otros (2003) señalaron que el entendimiento común de la relación "is-a" lo considera como "todo o nada", en el sentido que la relación es igualmente fuerte entre una clase y cualquiera de sus subclases y también en cada nivel de la jerarquía. Esta suposición ha sido considerada como una simplificación excesiva de la plasmación de las relaciones del mundo real por autores como Cohen y Murphy (1984) Sicilia y otros (2003), Cross (2004), Hu y otros (2007) o Garofalakis y otros (2009), apuntando que el descuido de este hecho puede provocar un problema de adecuación epistemológica (McCarthy & Hayes, 1968).

En la investigación llevada a cabo por Sicilia y otros (2003), proporcionaron un ejemplo extremo, suponiendo una jerarquía con inicio en la clase "mamífero", con subclases de "gato doméstico" y "primate", y "siamés" como una subclase de "gato doméstico". Siguiendo su ejemplo es fácil apreciar subjetivamente que el primer nivel de especialización representa un paso más grande que el segundo y que la distancia entre la categoría de mamíferos y el primate es algo más corta que su distancia al gato doméstico.

Para superar esta situación, se propuso asignar un peso o distancia relativa para representar el nivel de similitud entre la clase y su subclase. Como resultado de este trabajo, especificaron una forma de extender DAML + OIL para codificar relaciones de similitud dentro de archivos RDF e implementarlo en un prototipo web que contaba con un agente personalizado para la recomendación de productos, el cual se implementaba sobre las categorías de productos de UNSPSC ("United Nations Standard Products and Services Code", un sistema de cifrado que clasifica productos y servicios para fines comerciales a escala mundial) empleando esta relación de similitud para navegar por las categorías de productos. En otro artículo, García y Sicilia (2003) indican cómo podrían utilizarse las ontologías para dar apoyo directo a las tácticas de búsqueda descritas por Bates (1990) respecto a la relación "gen-spec", como se muestra en la tabla 1.

Tabla 1. Relación que presentan las tácticas de búsqueda de Bates con la relación "gen-spec", según García y Sicilia (2003)

Search tactics	Bates' description	Relationship with "gen-spec" relation
SUPER	To move upward hierarchically to a broader (superordinate) term.	Straightforward navigations through the generalization-specialization hierarchy of the ontology.
SUB	To move downward hierarchically to a more specific (subordinate) term.	
RELATE	To move sideways hierarchically to a coordinate term.	Involves "crawling" the generalization line of the concept up to its immediate covering concept (possibly more than one), and then deciding which specialization is closer to the original one.

Las interfaces de búsqueda que utilizan sugerencias de términos relacionados podrían hacer uso de esta relación, pero en algunos casos el número de términos "superordinados", "subordinados" y "coordinados", en el sentido descrito por Bates (1990), es muy elevado y no hay manera de calcular qué términos serían mejores sugerencias. Esto ocurrirá frecuentemente cuando la relación sea igualmente fuerte entre un "superordinado" y cualquiera de sus términos "subordinados" y también en todos los niveles de la jerarquía, por lo que proporcionar un grado de distancia entre términos podría ayudar en esta decisión.

De manera similar, enfoques de recuperación de información como la expansión de las consultas realizadas por usuarios empleando mecanismos de expansión de consultas basados en ontologías (Segura, García-Barriocanal, & Prieto, 2011), así como otras tácticas destinadas a aumentar el número de posibles resultados significativos, podrían beneficiarse de esta información.

Las distancias entre términos podrían ser fácilmente representadas, por ejemplo, utilizando diagramas con distancias diferentes entre los nodos. Esto supone que el uso de la distancia de especialización entre conceptos también podría ser empleada en técnicas de visualización KOS.

El uso de técnicas de visualización para representar, explorar y navegar a través de KOS es una herramienta de gran utilidad no sólo para facilitar tareas de búsqueda (Bates, 1989) sino también para fines educativos y de aprendizaje. Por otra parte, la construcción de este tipo de interfaces no es una tarea sencilla y las decisiones de diseño deberían considerarse teniendo en cuenta los objetivos específicos, los usuarios y los contextos de uso, así como las particularidades del KOS; de cara a potenciar la usabilidad y utilidad de las mismas.

La representación de un alto número de términos es uno de los retos más relevantes en cuanto a la visualización de KOS; enfoques como el descrito por Julien y otros (2013) tienen como objetivo facilitar la navegación ocultando ciertos términos. Contar con información sobre el grado de distancia entre estos conceptos coordinados podría apoyar la decisión sobre cómo mostrar u ocultar ciertos términos. Otro problema común reside en entender el contexto del concepto de esquema de clasificación y las relaciones que vinculan los términos (Martín-Moncunill, Sánchez-Alonso, Gaona-García, & Marianos, 2013), situación que se agrava en los casos en que los usuarios pueden apreciar subjetivamente que la distancia entre los diferentes niveles de especialización o entre los términos coordinados no es siempre la misma.

En consecuencia, la noción de distancia entre los conceptos relacionados con "gen-spec" podría ser una valiosa herramienta para facilitar las tareas de búsqueda y recuperación de información y también para simplificar y mejorar las técnicas de visualización de KOS. Este último aspecto se trata en profundidad más adelante en el presente capítulo, en la sección "Interfaces y visualización de KOS: tendencias y desafíos".

2.4 Extracción automática de palabras clave

Uno de las posibles barreras para la adopción de KOS dentro de la clasificación y localización de información pudiera residir en la propia necesidad y el esfuerzo necesario para vincular los recursos de información existentes a los términos de KOS. Es por ello que uno de los primeros estudios a plantear en el desarrollo de la investigación que comprende esta tesis es la evaluación de mecanismos automáticos de extracción de palabras clave, contrastándolo con los resultados obtenidos por humanos.

El uso de KOS como parte de los metadatos para describir recursos hace posible la existencia de colecciones con una descripción consistente y homogénea de los mismos, al contrario de lo que ocurre con aquellas que emplean frases o palabras sin clasificar (Zeng & Chan, 2004). Esto incentiva el uso de técnicas automatizadas de extracción de palabras clave que se producen en base a una serie concreta de términos o un KOS, como podrían ser el tesauro AGROVOC.

Las palabras clave (también se pueden denominar como "frases clave") podrían definirse como secuencias de palabras que caracterizan el tema de un documento y su contenido (Turney, 1999). Han demostrado ser herramientas valiosas en el contexto de la ciencia de la información y del conocimiento (Dahlberg, 2006), particularmente en tareas tales como anotación (Frank, Paynter, Witten, Gutwin, & Nevill-Manning, 1999) (Park, y otros) indexación (Hjørland B. , 2011), resumen (Al-Hashemi, 2010) o para averiguar el tema o "aboutness" (Hjørland B. , 2001) de un documento; lo que permite mejorar la recuperación y facilitar la categorización (Hjørland B. , 1998) y la navegación de la información (Jones & Paynter, 2001).

La asignación de palabras clave a los documentos (indexación) se puede hacer usando algún tipo de vocabulario controlado o elegir libremente frases representativas que aparecen en el cuerpo del documento, lo que se conoce como "indexación libre". Por otro lado, la extracción de palabras clave consiste en la selección de las frases características más importantes dentro del cuerpo de un documento que caracterizan su temática (Turney, Learning to extract keyphrases from text, 1999). Esto se puede lograr de forma automática mediante técnicas automáticas de extracción de palabras clave o manualmente por expertos humanos.

La extracción manual de palabras clave es una tarea muy lenta que requiere no sólo el conocimiento en el tema, sino también la experiencia en la tarea en sí misma. La información actual y el crecimiento de los datos (Lord, Macdonald, Lyon, & Giaretta, 2004) hacen que la extracción manual de palabras clave no sea adecuada para la gran mayoría de los casos, debido a razones de tiempo y recursos. En este contexto, la extracción automática de palabras clave basada en el tesauro (Lim, Wong, & Lim, 2013) parece ser una herramienta económica y rápida para afrontar el problema.

Entre las técnicas de extracción automática de palabras clave propuestas hasta la fecha destacan GenEx (Turney, 1999), KP-Miner (El-Beltagy, 2006) y KEA (Frank, Paynter, Witten, Gutwin, & Nevill-Manning, 1999) (Medelyan & Witten, 2005) (Medelyan, Milne, Legg, & Witten, 2009) por ser los más ampliamente empleados y referenciados (Lim, Wong, & Lim, 2013). Sin embargo, la información sobre la calidad y aplicabilidad de estas técnicas de extracción es limitada y desigual, siendo abordado por un número limitado de estudios.

La evaluación de la efectividad de este tipo de algoritmos se ha hecho típicamente utilizando medidas tipo precisión y exhaustividad (del inglés "precision / recall"), que comparan la lista de palabras clave generadas a las listas proporcionadas por el autor. La mayoría de estos estudios han sido limitados en términos de tamaño de la muestra, además, este enfoque sólo da una evaluación de la exactitud del sistema para hacer coincidir las palabras clave proporcionadas por el autor original (o curador de los datos), lo que plantea algunas preguntas importantes.

Por ejemplo, no es adecuado suponer directamente que las palabras clave originalmente seleccionadas fueron las más apropiadas, o que los expertos humanos obtendrían mejores resultados que los algoritmos al tratar de coincidir con las palabras clave del autor original. Además, desde el punto de vista de la aplicabilidad práctica, necesitamos conocer la calidad relativa de las palabras clave asignadas por el ser humano en comparación con las asignadas automáticamente.

2.4.1 Técnicas de extracción y mecanismos de evaluación de las mismas

El trabajo de realizado por Lim y otros (2013) proporciona una extensa revisión de las técnicas automáticas de extracción de palabras clave más relevantes, describiendo sus fortalezas y debilidades, clasificando las técnicas de extracción de palabras clave según dos criterios:

- El enfoque de aprendizaje (Turney, 2000) que puede ser "supervisado", "sin supervisión" o "no aprendizaje".
- El tipo de problema, que puede ser un problema de clasificación (Witten, y otros, 1999) o un problema de clasificación (Frantzi, Ananiadou, & Mima, 2000).

A continuación, se presentan las técnicas más relevantes en el ámbito que nos ocupa.

GenEx (Turney, 1999) es una técnica de extracción de palabras clave que combina dos módulos: "Extractor", el sistema automático de extracción de palabras clave y "Genitor", un sistema externo utilizado para calibrar el extractor de palabras clave, que utiliza 12 parámetros. El Extractor administra un proceso similar al árbol de decisión para realizar tanto la selección como la ponderación de los candidatos, sobre la base de tres atributos: frecuencia de término (TF), información de primera aparición y longitud de frase.

La frecuencia de término se utiliza como la puntuación base para el Extractor. GenEx no utiliza ningún atributo dependiente del dominio para propósitos de clasificación y ponderación. Por lo tanto, aunque requiere un largo período de formación inicial, no hay necesidad de volver a entrenar GenEx para cada nuevo dominio. La contribución más valiosa de GenEx es la capacidad de mantener su rendimiento en diferentes dominios (Frank, Paynter, Witten, Gutwin, & Nevill-Manning, 1999) (Lim, Wong, & Lim, 2013).

La técnica “multilayer perceptron technique” (MLP) (Sarkar, Nasipuri, & Ghose, 2010) fue concebida para abordar el problema de tener menos palabras clave generadas que el número solicitado, al tratar la extracción de frase clave como un problema de clasificación. Esta técnica supone que la extracción de palabras clave debe tratarse como un problema de ranking en lugar de un problema de clasificación. Los procesos MLP son similares a los de Extractor en GenEx (Turney, 1999) pero usan más atributos: TFxIDF (Salton, Wong, & Yang, 1975) y la combinación de longitud de frase y longitud de palabra. Además, con el fin de abordar la cuestión de la selección de un número insuficiente de palabras clave, MLP adjunta una lista de palabras “no clave” al final de la lista de palabras clave.

KP-Miner (El-Beltagy, 2006) presenta un enfoque de no-aprendizaje, basado en el ranking, lo que significa que no se necesita entrenamiento para el sistema. Esta técnica de extracción de palabras clave utiliza tres atributos: TFxIDF (Salton, Wong, & Yang, 1975), “First Occurrence Position” y un factor de impulso, que se centra en la selección de términos candidatos y en el proceso de ponderación. KP-Miner ha obtenido mejores resultados que otras técnicas que emplean un enfoque de aprendizaje, como GenEx y KEA (El-Beltagy & Rafea, 2009), pero ha sido criticado por la complejidad de su estructura, el uso de procesos innecesarios y la existencia de un sesgo en el cálculo de la frecuencia de los términos (Lim, Wong, & Lim, 2013). (Kumar & Srinathan, 2008) sugirieron mejorar KP-Miner introduciendo la Técnica de Filtración de “*N-gram*” en el proceso de ponderación, utilizando técnicas de compresión de datos LZ78 para generar la lista de palabras clave candidatas e ignorando IDF al elegir atributos de frase para ponderar.

KEA (del inglés, “Keyphrase Extraction Algorithm”), por (Frank, Paynter, Witten, Gutwin, & Nevill-Manning, 1999), se basa en un enfoque de aprendizaje supervisado en un contexto de problema de clasificación que emplea naïve Bayes como algoritmo de aprendizaje automático. KEA se centra en la selección de candidatos, la ponderación del término y el proceso de clasificación / clasificación, utilizando dos atributos para seleccionar las palabras clave del candidato: primera aparición y TFxIDF. El algoritmo de aprendizaje naïve Bayes permite que KEA requiera mucho menos tiempo de entrenamiento que GenEx, mientras que sigue funcionando aproximadamente al mismo nivel (Sarkar, Nasipuri, & Ghose, 2010). KEA permite mejorar los valores de precisión y exhaustividad en la extracción de palabras clave mediante la incorporación de la dependencia de dominio, basándose en la colección de corpus de dominio utilizados en los entrenamientos.

Como se mencionó anteriormente, la asignación de palabras clave a los documentos se puede realizar utilizando un vocabulario controlado o mediante "indexación libre". El estudio que ocupa el presente trabajo se centra en el primer enfoque, dado que el objetivo es proporcionar sistemas de información con un espacio de palabras clave uniforme para diferentes tareas, mediante el uso de las terminologías controladas que componen los KOS. Las técnicas de evaluación de los mecanismos de extracción de palabras clave se organizan en dos grandes grupos:

1. Evaluación automática: compara automáticamente los índices generados por la máquina con los originalmente asignados por el ser humano, estableciendo un estándar de oro. Este enfoque tiene algunos problemas, en particular que la elección de términos de índice ha sido criticada como subjetiva (Pouliquen, Steinberger, & Ignat, 2003).
2. Evaluación manual: los evaluadores humanos comparan el conjunto de índices generados por máquinas con el texto fuente, un proceso en el cual los evaluadores humanos usualmente realizan análisis cualitativos (Medelyan, Frank, & Witten, 2009).

2.4.2 Key Extraction Algorithm (KEA)

El experimento realizado en el presente trabajo está enfocado al uso de KOS de dominio específico, empleando AGROVOC como el KOS principal. KEA ha sido ampliamente utilizado junto con AGROVOC para extraer palabras clave de documentos en el dominio agrícola, por lo que resulta la herramienta más adecuada para la realización del estudio que se desarrolla en el capítulo 3.

Los resultados de las evaluaciones realizadas en experimentos anteriores son prometedores, pero no proporcionan toda la información de "aplicabilidad práctica" que persigue este trabajo. Además, el número de documentos analizados es reducido y las selecciones consideradas en experimentos anteriores parecían demasiado homogéneas.

Medelyan y Witten (2005) utilizaron KEA para extraer automáticamente términos de documentos relacionados con el dominio de la agricultura utilizando el tesauro AGROVOC desarrollado por la FAO como vocabulario controlado y como base de conocimiento para la correspondencia semántica, evaluando el algoritmo en un corpus de 200 documentos. Nguyen y Kan (2007) introdujeron la posibilidad de emplear KEA para capturar las posiciones de palabras clave en un documento con respecto a las secciones lógicas encontradas en un discurso científico, evaluando su algoritmo en un corpus de 120 documentos.

Se han realizado evaluaciones humanas de KEA, como el caso del experimento realizado por Medelyan y otros (2009), donde mostraron que las palabras clave asignadas automáticamente a los documentos tenían una precisión comparable a la de las asignaciones de los catalogadores humanos. En concreto, este experimento analizó la consistencia del etiquetado y

la frase clave del servicio CiteULike.org para la organización de citas académicas, utilizando un conjunto de 180 documentos indexados por 332 etiquetadores.

Jones y Painter (2001) evaluaron KEA basándose en evaluaciones subjetivas de la calidad y adecuación de palabras clave extraídas por expertos humanos. Se utilizaron como documentos de prueba un conjunto de seis artículos de las Actas de la Conferencia ACM sobre Factores Humanos de 1997, eliminando previamente las palabras clave de los autores. A continuación, se pidió a 28 sujetos que evaluaran cómo de adecuadamente representaban las palabras clave extraídas a cada artículo, mediante una escala de 0 a 10. Usando el Coeficiente de Kendall, el análisis demostró que había niveles de acuerdo significativos y a veces fuertes entre los sujetos al evaluar las palabras clave y que la mayoría de estas fueron valoradas positivamente.

Siguiendo este razonamiento, se ha seleccionado KEA como la técnica para la realización del experimento sobre las posibilidades del empleo de mecanismos de extracción automática de palabras clave presentado en esta tesis doctoral. Como se explicará en la sección destinada al mismo, el enfoque de evaluación humana realizado es diferente y proporciona mayor información sobre la aplicabilidad práctica, puesto que se realiza una triple comparación evaluando las coincidencias entre:

1. Las palabras clave originales, que fueron asignadas por curadores usando el vocabulario controlado de AGROVOC.
2. Las palabras clave extraídas por KEA
3. Las palabras clave extraídas por humanos expertos en el dominio.

Esta comparación incluye además una evaluación de expertos en el dominio, donde se realiza una valoración subjetiva de la adecuación de las palabras clave extraídas, sin conocer el mecanismo por el que han sido obtenidas.

En el momento de la realización de este experimento no se han encontrado otros previos que hubieran tomado este enfoque, por lo que pareció prometedor como una forma de generar evidencia adicional sobre los resultados de los algoritmos de extracción de palabras clave versus técnicas de extracción manual.

Por último, en la evaluación del rendimiento de KEA presentado en esta obra, el conjunto de recursos es considerablemente mayor que los empleados en los experimentos mencionados anteriormente con AGROVOC, siendo además un conjunto de recursos más heterogéneo y comprensivo, donde los documentos no proceden de la misma colección, de los mismos temas o de los mismos años.

2.5 Interfaces y visualización de KOS: tendencias y desafíos.

Las relaciones que se establecen entre los términos de un KOS pueden ser representadas visualmente, lo cual ha potenciado el uso de interfaces visuales para navegar a través de estos sistemas (Gaona García, Martín-Moncunill, Sánchez-Alonso, & García, 2014) y utilizarlos para formular consultas (García-Barriocanal & Sicilia-Urbán, 2003) seleccionando términos siguiendo las técnicas descritas por MJ Bates (1989).

El esfuerzo en cuanto a la investigación en el desarrollo de interfaces visuales que plasmen las relaciones y propiedades que se establecen en los sistemas de organización del conocimiento es notable y de largo recorrido. La mayoría de las propuestas enfatizan el permitir navegar de forma dinámica a través de las mismas y servir de soporte para localizar recursos mostrando claramente su contexto. Existen desarrollos muy recientes en esta línea que pretenden hacer uso de las posibilidades de realidad virtual y aumentada (Martín-Moncunill, Gaona-García, Gordillo-Orjuela, & Montenegro-Marin, 2016).

Como ya se ha comentado, los repositorios digitales constituyen uno de los principales casos de uso de los KOS y donde más se han integrado y ensayado distintas interfaces que permiten hacer uso de los mismos; principalmente ligados a tareas de búsqueda e indexación.

Los repositorios (Kennan & Wilson, 2006) y las librerías digitales (Isfandyari-Moghaddam & Bayat, 2008) son infraestructuras importantes, utilizadas para facilitar el acceso a colecciones de recursos digitales, las cuales son creadas por instituciones como museos, librerías o centros educativos, tales como Open Discovery Space (Nikolas, Sotiriou, Zervas, & Sampson, 2014), MERLOT (Burns & Schell, 2002), ARIADNE (Duval, y otros, 2001) y AGREGA2 (Sarasa, Canabal, & Sacristán, 2009).

La mayoría de bibliotecas digitales y repositorios, en particular los repositorios desarrollados por creadores y expertos, utilizan actualmente tipos de KOS para ofrecer métodos a través de los cuales gestionar, ordenar y organizar los distintos recursos que contienen, con el fin de optimizar el manejo de información y los procedimientos de búsqueda para sus recursos. El uso de repositorios y bibliotecas digitales se puede mejorar gracias a la implementación de KOS, para así ayudar a los usuarios a localizar los objetos de aprendizaje y mostrar información adicional, así como identificar relaciones, categorías y esferas de especialización, entre otras tareas.

Sin embargo, a día de hoy, el uso de interfaces visuales que permiten interactuar con KOS de forma consciente, no se han adoptado mucho más allá de propuestas teóricas, no se han extendido entre el público general, contradiciendo lo que a priori parece un enfoque acertado y útil. Esto lleva a preguntarse si realmente este tipo de interfaces carecen de utilidad real, si ésta es limitada frente a las clásicas interfaces textuales o si, por el contrario, esta utilidad existe y se debería seguir trabajando en el desarrollo de las mismas, contextualizándolas adecuadamente.

Un aspecto importante en la visualización de la información en cuanto a interfaces de búsqueda visuales, es la posibilidad que existe de incrementar la retentiva del usuario a través del uso de aspectos de diseño, tales como el color, el tamaño, la forma, la orientación, la posición, la organización y las relaciones de contenido.

Como ejemplo de esto se pueden mencionar los estudios de Baeza-Yates y Ribeiro-Nieto (2010), que indican que el propósito principal de los interfaces de búsqueda visual es reducir el trabajo mental del usuario; o el de Cao y otros (2010) que destacan la importancia de utilizar representaciones visuales en lugar de interfaces textuales para buscar recopilaciones de documentos a través de un proceso experimental.

Sin embargo, de acuerdo con Hearst (2009), aplicar visualización a la información textual es un desafío, especialmente cuando el propósito es mejorar la búsqueda de recopilaciones de texto. De hecho, existen multitud de estudios de usabilidad que indican que en la gran mayoría de los casos, la interfaz más adecuada para que un usuario realice búsquedas es un simple cuadro de texto (Nielsen J. , 2001) (Krug, 2005), por lo que la línea expuesta por autores como Baeza-Yates y Ribeiro-Nieto debería contextualizarse, en el sentido que indica la definición de usabilidad (ISO/IEC 9241): “Usabilidad es la eficacia, eficiencia y satisfacción con la que un producto permite alcanzar objetivos específicos a usuarios específicos en un contexto de uso específico”. Tal como expone Bevan (1999), la usabilidad depende no sólo del producto sino también del usuario, por lo que un producto no es en ningún caso intrínsecamente usable, sólo tendrá la capacidad de ser usado en un contexto particular y por usuarios particulares, esto es: la usabilidad no puede ser valorada estudiando un producto de manera aislada.

La representación de un alto número de términos es uno de los retos relevantes en cuanto a la visualización de KOS. Por ejemplo, en el tesoro AGROVOC hay más de 10 niveles de profundidad, en varios casos hay decenas (e incluso centenas) de términos coordinados y no todos los términos podrían categorizarse como específicos del dominio.

Este hecho dificulta enormemente su representación de una manera que sea útil para un usuario final que intentara navegar a través del mismo; no sólo debido a las limitaciones del tamaño de pantalla, sino también debido a la capacidad del usuario para evitar perderse mientras trataba de navegar a través de grandes cantidades de términos.

Enfoques como el descrito por Julien y otros (2013) tienen como objetivo facilitar la navegación ocultando ciertos términos. Como se mostrará en la sección dedicada a la distancia de especialización entre conceptos, dentro del capítulo 3, contar con información sobre el grado de distancia entre estos conceptos coordinados, podría apoyar la decisión sobre cómo mostrar u ocultar ciertos términos (Martín-Moncunill, García-Barriocanal, Sicilia, & Stracke, en prensa).

Otro problema común desde la perspectiva del usuario reside en entender el contexto del concepto de esquema de clasificación y las relaciones que vinculan los términos, especialmente cuando todos los términos y sus relaciones están representados de manera similar; esto es, los mismos colores, la misma forma, el mismo nivel de visualización, etc. Esta situación se agrava en los casos en que los usuarios pueden apreciar subjetivamente que la distancia entre los diferentes niveles de especialización o entre los términos coordinados no es siempre la misma (Sicilia, García, Aedo, & Díaz, 2003).

Desde un punto de vista teórico, las interfaces visuales pueden sacar partido de manera eficaz a las posibilidades creadas por el uso de KOS, pero desde un punto de vista práctico, las soluciones actualmente disponibles para los “usuarios finales” están muy lejos de las estrategias existentes, no existiendo una solución efectiva para los mismos. Por lo tanto, uno de los mayores desafíos que enfrentan los creadores de repositorios especializados de datos y bibliotecas digitales es la definición de métodos y estrategias para facilitar el acceso a recursos digitales.

Facilitar el acceso requiere definir apropiadamente las estrategias de búsqueda y los esquemas de organización del conocimiento, diseñar esquemas de metadatos y desarrollar mecanismos para comprobar la calidad de los mismos. El diseño de los interfaces visuales de búsqueda basado en las relaciones de los KOS es una tarea altamente compleja y uno de los principales retos está relacionado con la adaptabilidad de los KOS existentes, dado que la mayor parte de dichos sistemas no están concebidos para su uso en una interfaz e incluso ni siquiera para un contexto de uso o para un usuario particular.

El trabajo realizado en este ámbito se ha plasmado en una revisión sistemática de la literatura (Gaona-García, Martín-Moncunill, & Montenegro-Marin, 2017) centrada en las tendencias y desafíos de interfaces para el acceso a recursos en bibliotecas digitales y repositorios, de la cual se expondrán los aspectos más relevantes a continuación. Entre los estudios recogidos en la mencionada revisión sistemática de la literatura, se incluyen los distintos experimentos relativos a interfaces visuales de búsqueda que ha llevado a cabo el autor de la presente obra, los cuales se presentan en el capítulo 4.

El análisis realizado en la revisión sistemática de la literatura permitió definir una lista de tendencias y desafíos en el uso, diseño e integración de interfaces basadas en la visualización de la información (Card, Mackinlay, & Shneiderman, 1999) (Shneiderman & Plaisant, 2006) (Hearst, 2009) (Draper, Livnat, & Riesenfeld, 2009) (Graham & Kennedy, 2010) (Gleicher, y otros, 2011) (Gleicher, 2011), el diseño centrado en el usuario, partiendo de los principios de la interacción persona computador (Shneiderman & Plaisant, 2005) y el aprovechamiento del uso de metadatos (Palavitsinis, Manouselis, & Sanchez-Alonso, 2014) establecidos en los recursos digitales a través del uso de KOS que se empleen como método de clasificación para la localización de recursos (Ochoa, Klerkx, Vandeputte, & Duval, 2011).

Siguiendo el enfoque de revisión de la literatura propuesto por Kitchenham y otros (2009) las tres cuestiones fundamentales planteadas en el análisis fueron:

1. ¿Cuáles son los problemas más frecuentes en las interfaces de búsqueda para acceder a los recursos ofrecidos por bibliotecas y repositorios digitales?
2. ¿Cuáles son los desafíos para integrar interfaces visuales de búsqueda en bibliotecas y repositorios digitales?
3. ¿Cuáles son los aspectos y consideraciones que llevan a utilizar técnicas de visualización de la información para buscar recursos en bibliotecas y repositorios digitales?

En cuanto a la primera cuestión, destacar que existe una gran variedad de problemas en bibliotecas y repositorios digitales relativas a la facilidad para localizar y acceder a los recursos y la mayoría de ellos están relacionados a problemas de diseño en las interfaces. Autores como Hanisch y Straßer (2005) llegaron a afirmar que los repositorios carecen del respaldo necesario para desplegar recursos visuales e interactivos.

Multitud de estudios previos han descubierto que este tipo de interfaces no satisfacen de manera adecuada las necesidades del usuario enfrentándose, entre otras, a importantes barreras para encontrar recursos (Tenopir, 2003) (Khoo, Kusunoki, & MacDonald, 2012) o descubrir cuáles de los recursos encontrados no son los relevantes para su búsqueda (Nash, 2005). Autores como Hitchcock y otros (2003) o Tsakonas y Papatheodorou, (2008), inciden en que los repositorios deben permitir una visualización instantánea de los materiales disponibles.

Fuhr y otros (2007) aportaron una metodología para la clasificación de procesos de evaluación en las librerías digitales. El estudio muestra la relevancia de los usuarios reales en los procesos para obtener una mejor comunicación en la definición de una terminología común. Jovanović y otros (2007) mencionaron que el uso de ontologías en los recursos digitales incrementa la consistencia y la interoperabilidad de los metadatos.

Sin embargo, Gašević y Hatala (2006) señalan que el uso de ontologías también tiene desventajas para la accesibilidad a los recursos digitales porque los usuarios encuentran problemas para familiarizarse con una estructura ontológica predefinida y prefieren utilizar terminología común. Este aspecto también fue apreciado en los experimentos en cuanto a interfaces visuales reflejados en el capítulo 4, donde se muestra claramente cómo el uso de terminología común adecuada para los usuarios finales es un aspecto esencial en el diseño de KOS (Gašević, Djurić, & Devedzic, 2009).

En cuanto a las posibilidades de navegación, una limitación ampliamente conocida es conseguir la combinación óptima entre esta y los métodos de búsqueda (Hartson, Shivakumar, & Pérez-Quinones, 2004). A esto hay que sumar que cada tipo de KOS tiene ventajas y desventajas a la hora de representar el conocimiento que entrañan en una interfaz y no es posible desarrollar soluciones universales que satisfagan las necesidades básicas de los usuarios. De hecho, existe un consenso total entre los autores estudiados en la revisión sistemática de la literatura, en cuanto a que no existe un método genérico para la representación de conocimiento que pueda ser aplicado universalmente para estructurar datos en todos los casos.

Sobre la integración de interfaces, la gran mayoría de los autores como Aula y Käkki, (2005), Battenfield (1999) o Hearst (2009) sostienen que la simplicidad debe ser uno de los principios fundamentales para la construcción de las mismas. Cada KOS presenta especificidades en cuanto a la notación empleada para la representación del conocimiento, para determinar el área temática, el nivel de relación, los vínculos y las formas de asociación. Entre otros factores y como ya se indicó, esto dificulta la existencia de un método genérico para la representación de este conocimiento en una interfaz que pueda considerarse estándar. En cualquier caso, un principio de diseño a considerar de forma universal es el que el KOS debe ser consistente y realista cuando represente el área temática y las relaciones del mismo, para poder ser efectivo (Chrysafiady & Virvou, 2013). El uso de interfaces de visualización y KOS en repositorios y librerías digitales debería facilitar el acceso a los recursos, permitiendo al usuario hacerse una idea clara del alcance del repositorio y una mayor comprensión y asimilación de los recursos digitales y su contexto, dando lugar a una mejor comprensión del tema o área de conocimiento.

El conocimiento del dominio (Hjørland B. , 2002) de los términos incluidos en el KOS se puede considerar como un factor de gran relevancia a la hora de mejorar la experiencia de usuario de estas interfaces. Visto en sentido opuesto, un KOS desarrollado adecuadamente y plasmado de forma correcta será de utilidad para incrementar el conocimiento de los usuarios respecto al dominio. Respecto a este punto, un problema frecuente es que los KOS de dominio específico normalmente incluyen una considerable cantidad de términos que pueden considerarse genéricos, este hecho se analiza en profundidad en el capítulo 4.

La cuestión sobre estrategias para el uso de metadatos como criterio de búsqueda (Ruecker, Radzikowska, & Sinclair, 2011) en el ámbito de los repositorios y librerías digitales se ha planteado en una gran cantidad de estudios. Autores como Chirita y otros (2005), Hughes y Kamat (2005), Sánchez-Alonso & Sicilia (2005), Lytras y Sicilia (2007), Park (2009), Ochoa y Duval (2009) o Stawniak (2012); coinciden en que dichas estrategias facilitan los procesos de búsqueda para recursos digitales .

Los metadatos, independientemente de cómo se obtengan y estructuren, se consideran como un requerimiento esencial para la obtención de resultados de búsqueda de manera que se adecúen a las necesidades del usuario (Cechinel, Sánchez-Alonso, & Sicilia, 2009). Para esto es fundamental abarcar los problemas más frecuentes en la definición de metadatos relacionados con los recursos digitales, incluyendo la inconsistencia (Palavitsinis, Manouselis,

& Sanchez-Alonso, 2014), la redundancia (Ochoa & Duval, 2009) y la ambigüedad (Park J. R., 2009).

Por ejemplo, en un escenario educativo, las interfaces visuales de búsqueda deberían favorecer el proceso de encontrar material educativo, empleando sistemas para categorizar recursos por edad, temática, etc. La visualización de la información utilizando criterios de búsqueda que hacen uso de metadatos y KOS deberían permitir la rápida localización de los recursos e igualmente promover los procesos de exploración de recursos para aportar un valor añadido en el contexto educativo.

Como conclusión es fundamental destacar la importancia de mejorar no sólo la usabilidad de la interfaz, sino también la usabilidad de KOS de acuerdo con los perfiles del usuario, objetivos y uso contextual. Para esto, en la elaboración tanto del KOS como de la interfaz, se debería partir de una filosofía de diseño centrada en el usuario, que cubra todo el ciclo de desarrollo de las mismas, empleando técnicas que intenten optimizar la usabilidad.

Capítulo 3: Mecanismos de evaluación y análisis de los factores limitantes observados

“The sciences, each straining in its own direction, have hitherto harmed us little; but some day the piecing together of dissociated knowledge will open up such terrifying vistas of reality, and of our frightful position therein, that we shall either go mad from the revelation or flee from the deadly light into the peace and safety of a new dark age.”

(Extracto de “The Call of Cthulhu”, de H. P. Lovecraft, 1928)

3.1 Introducción

En este capítulo se presentan los experimentos realizados que tratan de dar respuesta a las preguntas planteadas en la introducción y realizar las aportaciones definidas en los objetivos.

La causa más inmediata que puede intuirse en cuanto al desuso de los KOS es que las interfaces a través de las cuales se les debería de dar uso no son adecuadas. Es por ello que la motivación de esta tesis se inició en un estudio de usabilidad de interfaces y, a lo largo de toda la investigación plasmada en esta tesis, se han continuado realizando estudios que analizan interfaces que hacen uso de KOS y que aparecen recogidas en el capítulo 4.

En el estudio sobre las interfaces de Organic.Edunet (Martín-Moncunill, 2012) se identifican múltiples problemas de usabilidad ligados puramente a las interfaces, pero adicionalmente se localizan varios problemas que estarían por encima de la propia interfaz ligados a la apreciación de la utilidad, la comprensión y usabilidad del propio KOS. Esto se vuelve a apreciar en otros dos estudios relacionados con interfaces (Martín-Moncunill, Sánchez-Alonso, Gaona-García, & Marianos, 2013) y visualización (Gaona García, Martín-Moncunill, Sánchez-Alonso, & García, 2014) de KOS.

Este hecho motiva el planteamiento de cuestiones en cuanto a la utilidad, usabilidad e incluso la viabilidad de uso de los KOS. En este sentido, la primera cuestión que se plantea es, dado el enorme esfuerzo que supone la catalogación de recursos con KOS ¿es la imposibilidad de etiquetarlos adecuadamente un impedimento para la viabilidad de su adopción? En base a esto se plantea un estudio sobre los mecanismos de extracción automática de palabras clave, como herramienta para la catalogación automática de recursos (Martín-Moncunill, García-Barriocanal, Sicilia, & Sanchez-Alonso, Evaluating the Practical Applicability of Thesaurus-Based Keyphrase Extraction in the Agricultural Domain: Insights from the VOA3R Project, 2015).

La investigación en torno a la catalogación automática de recursos se realizó empleando AGROVOC, un KOS específico del dominio de la agricultura. Aunque existen KOS no específicos de dominio, los KOS mayoritariamente constituyen una forma de representación del conocimiento que tiene como objetivo organizar una terminología para un propósito particular

dentro de un dominio concreto (Hjørland B. , 2001). Durante la realización del experimento de catalogación automática de recursos se observó que había conjuntos de términos que no eran específicos de dominio, problema que se hizo visible en el estudio realizado en Organic.Edunet (Martín-Moncunill, 2012) y que podría limitar en cierta forma el uso de un KOS. Por ello, el siguiente experimento (Martín-Moncunill, Sicilia, García-Barriocanal, & Sanchez-Alonso, 2015) se centró en la evaluación de la especificidad de dominio en KOS.

Paralelamente se siguió trabajando la línea de evaluación de interfaces cuyos resultados, resumidos en el capítulo 4, continuaron apuntando hacia la problemática vista en el primer experimento (Martín-Moncunill, Sánchez-Alonso, Gaona-García, & Marianos, 2013) y la falta de atención a la hora de integrar el KOS con los requisitos de la interfaz en base a los objetivos y contexto de uso específico de los usuarios a los que estaban destinadas. La investigación realizada en cuanto a interfaces y visualización de KOS hasta ese momento quedó plasmada en un análisis sistemático de la literatura (Gaona-García, Martín-Moncunill, & Montenegro-Marin, 2017) relativo a las tendencias y desafíos de las mismas en el contexto de los repositorios y bibliotecas digitales.

La realización del experimento sobre especificidad de dominio, dejó también patente los problemas observados en cuanto a la comprensión de la jerarquía del KOS, la necesidad de revisar la coherencia de la misma y la posible utilidad que tendría emplear un valor de distancia de especialización entre conceptos (Sicilia, García, Aedo, & Díaz, 2003). Concretamente, durante el experimento, varios usuarios hicieron notar el hecho de las diferentes distancias en el sentido descrito por (Sicilia, García, Aedo, & Díaz, 2003) entre un nodo padre con los nodos hijos, llegando algunos a catalogarlo incluso como un error en el KOS. Este hecho reforzó lo detectado durante la evaluación del portal Organic.Edunet (Martín-Moncunill, 2012) y motivó la realización de un experimento sobre la evaluación de la distancia de especialización entre conceptos (Martín-Moncunill, García-Barriocanal, Sicilia, & Stracke, en prensa), siendo este el último experimento recogido en este capítulo.

3.2 Evaluación de la aplicabilidad práctica de la extracción automática de palabras clave

“For the dreams and secrets of our heart may be spoken, but words are poor handmaidens. Words can never fully say what we want them to say, for they fumble, stammer, and break the best porcelain. The best one can hope for is to find along the way someone to share the path, content to walk in silence, for the heart communes best when it does not try to speak.”

(Extracto de “Dragons of a lost star”, colección Dragonlance, M. Weis y T. Hickman, 2001)

3.2.1 Motivación

El uso de KOS como parte de los metadatos para describir recursos hace posible la existencia de colecciones con una descripción consistente y homogénea de los mismos (Zeng & Chan, 2004). Partiendo del deseo de emplear un KOS existente que se haya considerado como adecuado, el primer problema en la integración del mismo será vincular los recursos existentes a los clasificadores del KOS.

Por otro lado, los repositorios que ya emplean algún tipo de KOS para catalogar sus recursos se pueden enfrentar a problemas cuando intentan incluir o adecuarse a un perfil de metadatos. Los esquemas de metadatos empleados para describir los recursos de información suelen incluir clasificaciones o descripciones que utilizan algún tipo KOS concreto, ya sea un vocabulario controlado, un tesoro, un sistema de clasificación, una ontología, etc. Esto es un hecho habitual en las grandes colecciones digitales que contienen colecciones de fuentes que utilizan diferentes KOS y donde las descripciones deben alinearse.

El mapeo KOS puede proporcionar una solución al problema al que se enfrentan los grandes repositorios de recursos (Anibaldi, Jaques, Celli, Stellato, & Keizer, 2013) cuando los distintos KOS están correctamente mapeados; sin embargo, en la gran mayoría de los casos estos mapeos no están disponibles (Chen & Chen, 2012) o su cobertura se limita a pequeñas partes del KOS original. Esto incentiva el uso de técnicas automatizadas de extracción de palabras clave que producen términos en la terminología común de elección, proporcionando así la descripción homogénea necesaria.

Los algoritmos de catalogación automática de recursos podrían constituir una solución viable a este problema, pero los mecanismos de evaluación empleados hasta la fecha se basan en comprobar la exactitud del algoritmo para hacer coincidir exactamente las palabras clave proporcionadas por el autor original. Este indicador no refleja claramente la calidad y la aplicabilidad práctica de estas técnicas en comparación a los resultados que podrían obtener catalogadores humanos.

La investigación plasmada en esta sección pretende proporcionar una respuesta más certera a si los algoritmos de catalogación automática de recursos constituyen una solución viable, proporcionando un nivel de calidad adecuado, frente al problema que puede constituir la catalogación manual, esto es, la pregunta de investigación número 2 enunciada en el capítulo de introducción.

3.2.2 Introducción

Hasta la fecha se han propuesto una serie de técnicas para la extracción de palabras clave; GenEx (Turney 1999), KP-Miner (El-Beltagy 2006) y KEA (Medelyan y Witten, 2005) son algunos de los más populares y más referenciados (Lim y otros 2013). Sin embargo, la información sobre la calidad y aplicabilidad de estas técnicas de extracción de palabras clave es limitada.

La evaluación de la efectividad de este tipo de algoritmos se ha hecho típicamente utilizando medidas tipo precisión y exhaustividad (del inglés “precision / recall”), que comparan la lista de palabras clave generadas y las listas proporcionadas por el autor.

El hecho de que este enfoque sólo proporcione una evaluación de la exactitud del sistema para hacer coincidir las palabras clave proporcionadas por el autor original, hace que los responsables de las colecciones de recursos no estén seguros sobre la calidad y la aplicabilidad práctica de estas técnicas a sus servicios integrados, planteándose preguntas como:

- ¿Fueron las palabras clave originalmente seleccionadas por el autor las más apropiadas?
- ¿Humanos expertos en el campo obtendrían mejores resultados que los algoritmos al tratar de coincidir con las palabras clave del autor original?
- ¿Cuál es la calidad relativa de las palabras clave asignadas por el ser humano en comparación con las asignadas de forma automática?

Este experimento intenta llenar el vacío en cuanto a estas cuestiones al proporcionar evidencias adicionales sobre los resultados de los algoritmos de extracción de palabras clave aplicados en el contexto del proyecto VOA3R (Stracke, Sicilia, Helmstedt, Martín-Moncunill, & Protonotarios, 2013).

"VOA3R - Virtual Open Access Agriculture & Aquaculture Repository: Sharing Scientific and Scholarly Research related to Agriculture, Food, and Environment., lanzado en junio de 2010 y finalizado en 2013 con la valoración de “Excellent Progress” fue un proyecto financiado por la Comisión Europea bajo el programa PSP de CIP. El objetivo general de VOA3R es mejorar la difusión de los resultados de la investigación agrícola y acuícola europea utilizando un enfoque innovador para compartir resultados de investigación de acceso abierto. VOA3R conecta varios

sistemas de publicación en el ámbito de la agricultura y la acuicultura, incluyendo AGRIS bajo una estricta política de acceso abierto.

VOA3R persigue implementar un servicio integrado avanzado y centrado en la comunidad para la recuperación de contenidos abiertos. Para ello la plataforma VOA3R emplea tecnologías maduras en el área de los metadatos y semántica, como el perfil de aplicación AGRIS y el vocabulario controlado AGROVOC.

El buen funcionamiento de los servicios VOA3R requiere no sólo la integración de metadatos provenientes de distintas colecciones y catalogados con distintos KOS, sino también del enriquecimiento de los mismos; esto motivó la realización de un estudio piloto exploratorio para proporcionar ideas sobre la evaluación de la aplicabilidad práctica de las técnicas de extracción de palabras clave para este propósito.

Se utilizó el *framework* de código abierto KEA (Witten, y otros, 1999) para extraer palabras clave de grandes muestras de la colección AGRIS, una colección temática integral de información científica en el campo de la agricultura; utilizando el vocabulario de AGROVOC (Sini, Lauser, Salokhe, Keizer, & Katz, 2008); un vocabulario controlado que abarca todas las áreas de interés para la FAO. Como se ha visto en el capítulo destinado a la revisión de la literatura, KEA ha sido empleado en múltiples ocasiones junto a AGROVOC para extraer palabras clave de documentos en el dominio agrícola, existiendo varios estudios sobre su eficacia, por lo que resultaba la herramienta más adecuada para el presente estudio.

Los resultados muestran que el rendimiento de KEA al intentar obtener las palabras clave originales fue similar al observado en experimentos anteriores realizados y ligeramente mejor que el promedio humano. Los evaluadores, que participaron en un estudio ciego (no sabían si las palabras clave extraídas las había obtenido otro evaluador experto, el autor original o se habían obtenido de manera automática), consideraron que la calidad de las palabras clave extraídas, estaban además en un nivel similar al desempeño humano.

3.2.3 Materiales y método

La evaluación aquí descrita se realizó utilizando el conjunto de documentos de la colección AGRIS de la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO), que fue seleccionado para ser integrado en VOA3R y el vocabulario de AGROVOC.

El proceso de recolección y selección de datos fue complejo, involucrando diferentes etapas, incluyendo la elaboración de una herramienta de software para analizar y clasificar los registros.

A lo largo de esta sección se presentará la colección AGRIS y proporcionará información sobre el número de documentos clasificados por año y por tipo. A continuación, se analizará el conjunto de documentos AGRIS seleccionados por la FAO y el consorcio VOA3R para ser

integrados en la colección VOA3R, con el fin de seleccionar correctamente y reunir un conjunto adecuado para el experimento.

Los principales pasos del proceso, detallado más adelante, fueron:

1. Extracción del número de palabras clave de cada documento empleando AGROVOC. El análisis de estos datos permitió iniciar el proceso de selección según el número de palabras clave.
2. Localización de recursos en inglés que puedan ser procesados por el software de KEA. De este análisis surgió el hecho de que sólo alrededor del 3% de la colección completa de AGRIS tenía un enlace de texto completo disponible en sus metadatos. Además, no todos estos enlaces apuntaban a documentos de texto completo adecuados, sino a resúmenes o documentos escaneados (formato de imagen) que no eran adecuados para ser procesados por KEA.
3. Elaboración de un conjunto de documentos lo más variado posible de cara a la evaluación. Como parte de esta tarea, se tuvo que trabajar con el Número de Recurso AGRIS (ARN) que contiene información sobre la procedencia del recurso: el país, el año de la inclusión en la colección y el código del subcentro.
4. A fin de garantizar la adecuación de los puntos anteriores, se seleccionó manualmente una muestra de 2000 documentos de texto completo.
5. Finalmente se realizó una selección de cinco documentos para comparar palabras clave extraídas automáticamente y generadas por expertos.

La colección AGRIS

AGRIS es una colección de más de 7,6 millones de registros bibliográficos sobre temas de ciencia y tecnología agrícola. Es uno de los sistemas mundiales de información más importantes en el ámbito agrícola, sirviendo más de un millón de páginas al mes y más de doscientos cincuenta mil usuarios accediendo al sistema cada mes.

Sus registros cubren los diversos campos de la agricultura en un sentido amplio, incluyendo la silvicultura, la ganadería, las ciencias acuáticas y la pesca y la nutrición humana, desde 1975 hasta la actualidad. Se pueden encontrar metadatos sobre varios tipos de documentos y recursos, incluidos informes científicos y técnicos, publicaciones gubernamentales, tesis y documentos de conferencias, entre otros.

Esto hace que AGRIS sea una base de datos muy completa, como se puede ver en la figura 3, que muestra la distribución de recursos de AGRIS por año (elemento 4.5.2 "Fecha de publicación" del AGRIS AP) y la figura 4, que muestra la distribución de los recursos de AGRIS por tipo de publicación (elemento 4.9 "Tipo" del perfil de aplicación de AGRIS).

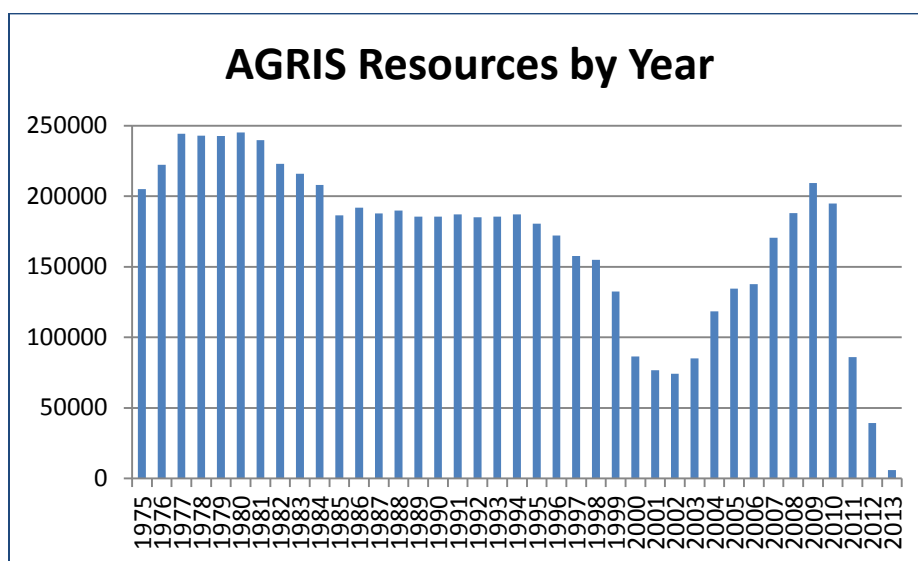


Figura 3. Recursos de AGRIS por año hasta 2013 (extraído en Noviembre de 2013)

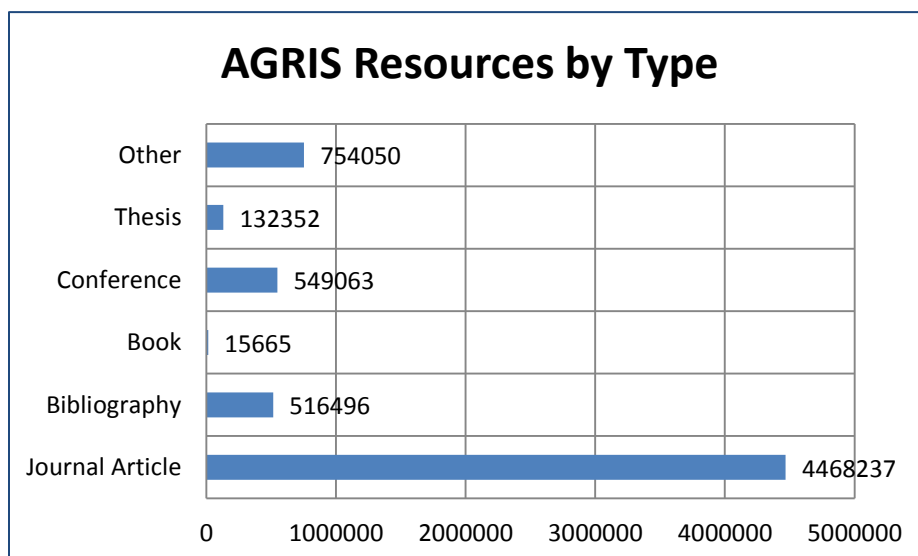


Figura 4. Recursos de AGRIS por tipo de recurso hasta 2013 (extraído en Noviembre de 2013)

El Perfil de Aplicación AGRIS (AGRIS AP) es el esquema estándar de metadatos creado por la FAO para definir los recursos enumerados en AGRIS. El elemento AGRIS AP "4.6.3 Subject Thesaurus" se utiliza para proporcionar palabras clave que describan el contenido del recurso e indiquen los descriptores que forman parte de un vocabulario controlado. La Figura 5 muestra un ejemplo de los metadatos AGRIS AP.

```
<ags:subjectThesaurus xml:lang="en"  
scheme="ags:AGROVOC">Mentha pulegium</ags:subjectThesaurus>  
<ags:subjectThesaurus xml:lang="en"  
scheme="ags:AGROVOC">chemical  
composition</ags:subjectThesaurus>  
<ags:subjectThesaurus xml:lang="en"  
scheme="ags:AGROVOC">essential oils</ags:subjectThesaurus>  
<ags:subjectThesaurus xml:lang="en"  
scheme="ags:AGROVOC">antimicrobial  
properties</ags:subjectThesaurus>
```

Figura 5. Un fragmento de los metadatos de un recurso descritos siguiendo el perfil de aplicación de AGRIS (AGRIS AP)

AGROVOC es el tesoro corporativo de la FAO. Abarca temas relacionados con los intereses de esta organización, incluyendo agricultura, silvicultura, pesca, medio ambiente y dominios relacionados. Se desarrolló por primera vez en los años ochenta para estandarizar el proceso de indexación de la base de datos de AGRIS con el fin de hacer la búsqueda más sencilla y más eficiente y guiar al usuario hacia los recursos más relevantes. Hoy AGROVOC contiene más de 40.000 conceptos organizados en una jerarquía y cada concepto puede tener etiquetas en hasta 22 idiomas. Es ampliamente utilizado por investigadores, bibliotecarios y administradores de información para indexar, recuperar y organizar datos en sistemas de información agrícola y páginas web en todo el mundo (Šimek, Vaněk, Jarolímek, Stočes, & Vogeltanzová, 2013). Hoy en día, AGROVOC se expresa en SKOS ("Simple Knowledge Organization System"), publicado como datos enlazados y alineado con diez otros KOS multilingües (Anibaldi, Jaques, Celli, Stellato, & Keizer, 2013).

Como puede verse en el análisis anterior, el tamaño, el número de recursos y la distribución de AGRIS representan una colección que abarca con amplitud las distintas temáticas del campo de la agricultura. Del mismo modo, la cobertura de vocabulario de AGROVOC, el número de términos y la jerarquía representa un vocabulario muy completo. Esto hace que AGRIS y AGROVOC sean ideales para el objeto del estudio.

Selección de la muestra

Como se describe en la introducción, la necesidad de integrar y enriquecer metadatos para la plataforma VOA3R brindó la oportunidad de experimentar y analizar la aplicabilidad práctica del servicio de indexación automática KEA en un contexto real, con una colección de documentos significativamente mayor que la utilizada en los experimentos relacionados con AGROVOC mencionados en el análisis de la literatura.

Entre sus colecciones, VOA3R contenía una selección de documentos procedentes de AGROVOC. La selección de estos documentos se realizó en el contexto del paquete de trabajo del denominado "*Content Providers*" (proveedores de contenido) del proyecto VOA3R a cargo de los expertos del consorcio y en colaboración directa con los expertos de la FAO.

El conjunto de documentos elegido como muestra para este experimento se seleccionó de la colección AGRIS en VOA3R, una colección de casi 3 millones de artículos proporcionados por la FAO para integrarse en la plataforma VOA3R. La selección de una muestra de esta colección requirió la implementación de un pequeño software para extraer la información necesaria y preparar todos los materiales necesarios para el experimento, con el fin de asegurar la aplicabilidad de los resultados a la muestra. Este software permitió recopilar datos estadísticos valiosos, analizar la colección AGRIS seleccionada para VOA3R y crear un archivo que contenía las palabras clave para cada recurso. Por otro lado, teniendo en cuenta que los registros bibliográficos de AGRIS son creados manualmente por los catalogadores y no siempre se rellenan de forma adecuada, este paso resultó también útil para clasificar los recursos y establecer el entorno necesario para ejecutar correctamente KEA.

Ejemplo de esto es que frecuente los registros de metadatos no incluyen una referencia al texto completo. De hecho, usando este software, se pudo apreciar que sólo alrededor del 3% de la colección completa de AGRIS contaba con un enlace al texto completo válido en sus metadatos. Este hecho obligó a ampliar el sistema para localizar los textos completos a través de Google Scholar (Harzing & Van der Wal, 2007), utilizando la información contenida en los registros de AGRIS.

El número total de registros de colección de VOA3R-AGRIS fue de 2.939.982. Se evaluó el número de palabras clave AGROVOC en inglés, que varió de 0 a 77 palabras clave por documento. La Figura 6 muestra la distribución general de palabras clave por documento, comenzando con documentos con una palabra clave.

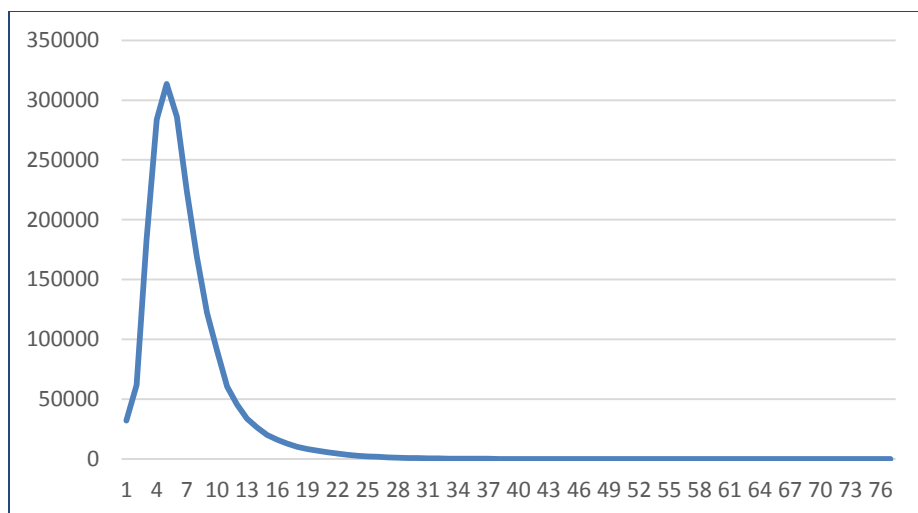


Figura 6. Distribución de palabras clave en inglés de la colección VOA3R-AGRIS

En la distribución anterior, la moda de palabras clave fue de 5 y la media de 6.9 palabras clave por recurso. La figura 7 proporciona más detalles en el intervalo de una a diez palabras clave.

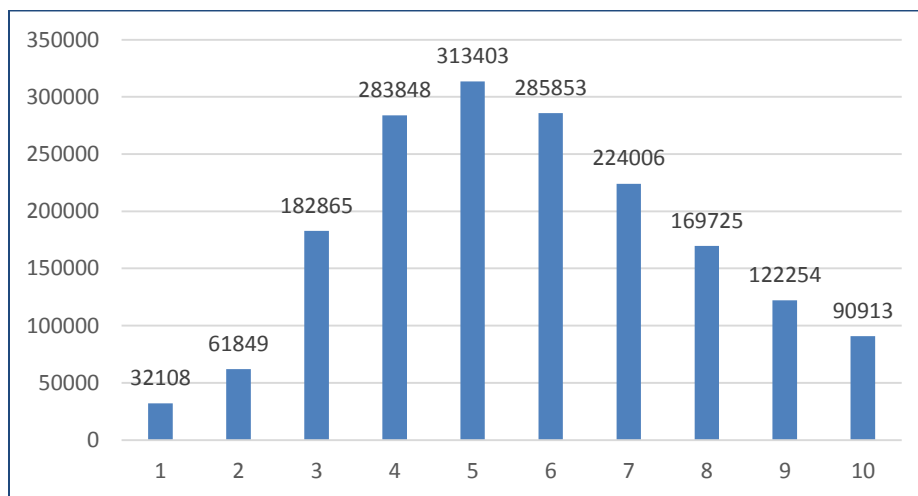


Figura 7. Detalle del número de recursos de la colección VOA3R-AGRIS según su número de palabras clave, en el intervalo de 1 a 10 palabras clave.

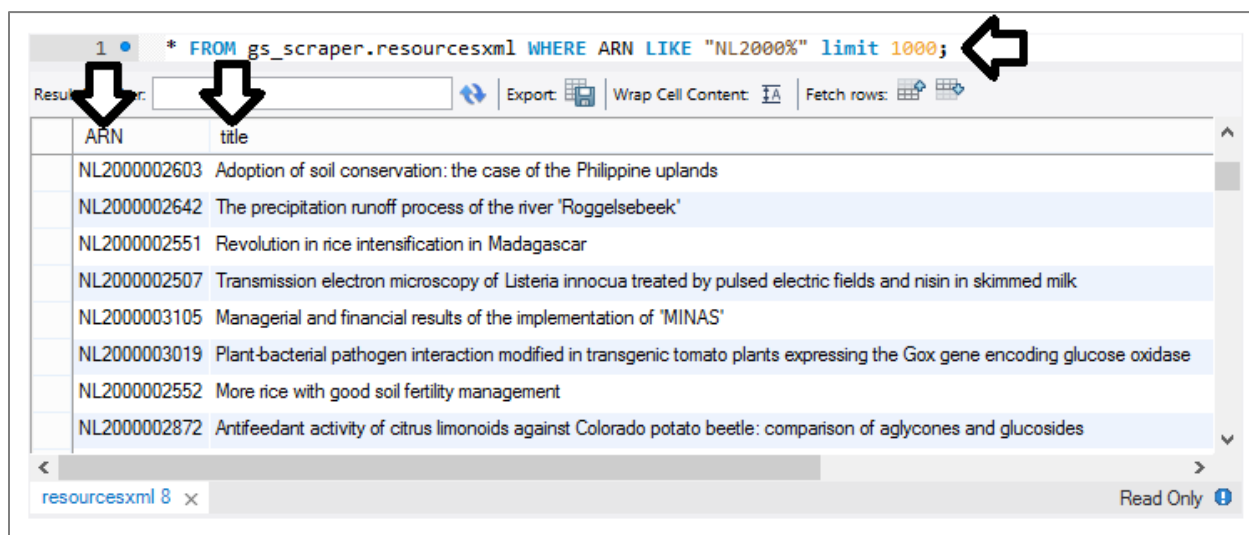
Los documentos sin palabras clave (905,451 o 31%) fueron descartados para el estudio. Con el fin de garantizar la aplicabilidad de los resultados a la colección se procuró encontrar un conjunto de documentos tan completo como fuera posible con respecto no sólo al número de palabras clave asignadas originalmente, sino también a atributos como el año de publicación, país o tema.

Para facilitar esto, se construyó una base de datos relacional con la información del ARN (AGRIS Resource Number) y el título de cada recurso (figura 6). ARN es un identificador

alfanumérico único para los recursos de AGRIS, cuya estructura proporciona directamente información sobre el país del que proviene el recurso, el año de inclusión, el código del subcentro y un número de serie, como se muestra en la figura 7.

Los documentos fueron descargados y revisados a mano para comprobar la idoneidad de cada uno, evitando los documentos escaneados directamente (que no contienen texto, sólo imágenes, un problema común con documentos antiguos), documentos que contienen sólo un resumen y otros documentos con problemas similares, buscando garantizar de la mejor forma posible la validez del experimento.

Por último, dado que KEA procesa documentos TXT, todos los artículos descargados (habitualmente en PDF) se convirtieron a TXT (UTF-8). De esta manera, se seleccionaron manualmente una muestra de 2000 documentos de texto completo que eran adecuados para el experimento, de acuerdo con los criterios previamente expuestos. El corpus de entrenamiento contenía 1200 documentos.



1 • * FROM gs_scraper.resourcesxml WHERE ARN LIKE "NL2000%" limit 1000;

Result: Export: Wrap Cell Content: Fetch rows:

ARN	title
NL2000002603	Adoption of soil conservation: the case of the Philippine uplands
NL2000002642	The precipitation runoff process of the river 'Roggelsebeek'
NL2000002551	Revolution in rice intensification in Madagascar
NL2000002507	Transmission electron microscopy of <i>Listeria innocua</i> treated by pulsed electric fields and nisin in skimmed milk
NL2000003105	Managerial and financial results of the implementation of 'MINAS'
NL2000003019	Plant-bacterial pathogen interaction modified in transgenic tomato plants expressing the Gox gene encoding glucose oxidase
NL2000002552	More rice with good soil fertility management
NL2000002872	Antifeedant activity of citrus limonoids against Colorado potato beetle: comparison of aglycones and glucosides

resourcesxml 8 x Read Only

Figura 8. La base de datos relacional creada, mostrando información sobre los recursos provenientes de los Países Bajos (NL) en el año 2000.

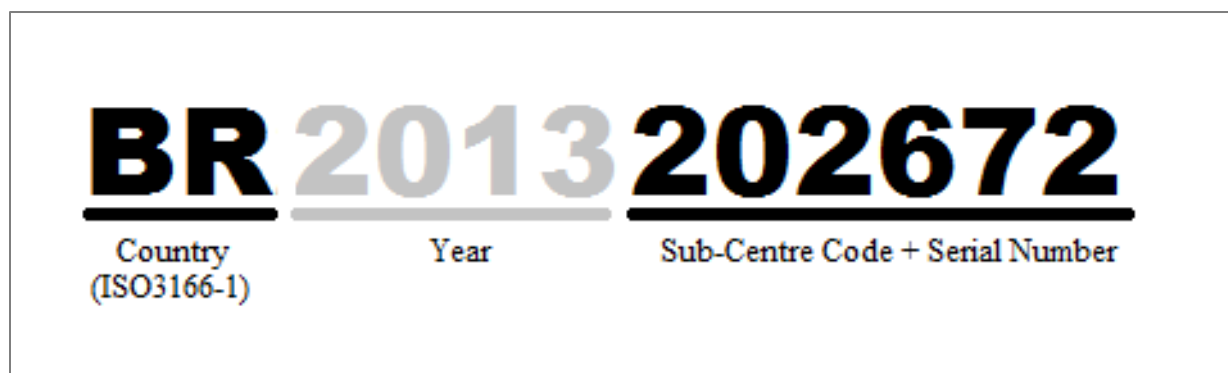


Figura 9. Información contenida en un número de registro ARN

Medición de la eficacia de KEA

Se evaluó la efectividad de KEA usando el criterio de precisión y exhaustividad ("*precisión / recall*") adoptado por van Rijsbergen (1979), que consiste en comparar los conjuntos de palabras clave del algoritmo con palabras clave asignadas manualmente. El número de palabras clave coincidentes se expresa como una proporción de todas las palabras clave extraídas (*Precision* -"P") y del número de palabras asignadas originalmente (manualmente) (*Recall* - "R") para cada documento por separado.

$$P = \frac{\#correct_extracted_keyphrases}{\#all_extracted_keyphrases} \quad R = \frac{\#correct_extracted_keyphrases}{\#manually_assigned_keyphrases}$$

Para la evaluación de la indexación automática y la evaluación de indexación humana, se define una palabra clave KEA como correcta si es una coincidencia exacta para la palabra clave original. La precisión se calcula dividiendo el número de palabras clave correctamente extraídas por el número total de palabras clave extraídas. La exhaustividad se calcula dividiendo el número de palabras clave extraídas correctamente por el número de palabras asignadas originalmente. Entonces, una combinación equilibrada de los dos se expresa por la medida F, como sigue:

$$F - measure = \frac{2PR}{P + R}$$

Comparando las palabras clave generadas por expertos con las extraídas automáticamente

La segunda parte del experimento comparó los resultados de KEA con una muestra de palabras clave extraídas por expertos humanos. Todos los expertos invitados a participar en el experimento estaban familiarizados con los dominios de la colección VOA3R-AGRIS, por lo que los documentos seleccionados para esta parte del experimento se extrajeron del mismo corpus utilizado para la evaluación automática.

La extracción de palabras clave utilizando KOS es una tarea que realmente requiere una gran cantidad de tiempo, experiencia y conocimiento en el campo. Por este motivo se seleccionaron artículos de investigación en el orden de 6 a 12 páginas, comprobando manualmente que eran adecuados para el experimento, con el fin de asegurar que todos los expertos invitados tuvieran suficiente conocimiento en el tema como para extraer palabras clave, que los artículos estuvieran completos y que no hubiera ningún otro factor que pudiera afectar la evaluación, incluyendo una comprobación adicional para omitir referencias directas a las palabras clave originales. La muestra final de cinco documentos se seleccionó aleatoriamente de documentos que se adecuaban a las características mencionadas.

A seis evaluadores humanos se les pidió asignar manualmente palabras clave empleando los términos AGROVOC. Se recomendó a los expertos que extrajeran un mínimo de seis palabras clave, teniendo en cuenta que, como se muestra en la figura 6, la moda de palabras clave originalmente asignadas en la colección VOA3R-AGRIS fue de 5 y la media fue de 6,9. No se fijó ningún tiempo máximo para que los expertos completaran estas tareas. Los expertos indicaron un tiempo de 20 a 40 minutos para la extracción de palabras clave de cada artículo.

Para profundizar en la aplicabilidad práctica de los resultados, se tuvo en cuenta que la elección de los términos está sujeta a un grado de subjetividad; al aceptar sólo coincidencias exactas con las palabras clave asignadas originalmente, el análisis resulta incompleto para evaluar su aplicabilidad práctica, ya que, en algunos casos, términos semánticamente cercanos o con similitud morfológica, también podrían ser útiles desde un punto de vista práctico.

Para evaluar esta aplicabilidad práctica, se solicitó a todos los expertos humanos que evaluaran la calidad de las palabras clave que se seleccionaron para clasificar los artículos de investigación, empleando una escala Likert de 5 puntos donde 1 indicaba "Representación completamente incorrecta" y 5 "Perfectamente bien representados". Para evitar sesgos en la evaluación, los sujetos en el estudio no sabían cómo se generaron estas palabras clave: si por otro experto humano, el autor original o automáticamente.

De esta forma, como complemento al enfoque utilizado por Jones y Painter (2011), en nuestro estudio cada evaluador humano evaluó no sólo las palabras clave asignadas automáticamente por KEA, sino también las asignadas manualmente por sus colegas, impidiendo a los evaluadores conocer cómo se asignaron las palabras clave. Este tipo de evaluación ciega era una necesidad que aparece recogida en la revisión de la literatura realizada por El-Haj y otros (2013).

3.2.4 Discusión de los resultados

Evaluación de la indexación automática

Como se describió anteriormente, durante la evaluación de indexación automática, KEA generó una serie de palabras clave que se compararon con las asignadas manualmente por expertos humanos. La tabla 2 muestra los resultados en términos de precisión y exhaustividad.

Tabla 2. Resultados para la evaluación automática de la indexación, en términos de precisión y exhaustividad utilizando la muestra seleccionada de 2000 documentos (el corpus de entrenamiento tenía 1200 documentos).

Precision	Recall	F-measure
0.18	0.20	0.19

Los resultados obtenidos fueron similares a los obtenidos en experimentos previos que evalúan la indexación automática basada en KOS, como los revisados por Lim y otros (2013). La tabla 3 compara los resultados calculados para la precisión, la exhaustividad y la medida F de acuerdo con esta revisión de la literatura.

Tabla 3. Resultados calculados para la evaluación automática de la indexación, en términos de precisión, exhaustividad y medida F. Derivado de la Tabla 1 en Lim et al. (2013).

Sources	Average Precision, Recall and F-measure		
	Precision	Recall	F-measure
Turney (2002)	0.18	N/A	N/A
Frank et al. (2005)	0.23	N/A	N/A
El-Beltagy et al. (2008)	0.16	0.33	0.21
Kumar et al. (2008)	0.21	0.31	0.25
Sarkar et al. (2010)	0.13	0.44	0.20
Average	0.182	0.36	0.22

Los resultados del presente experimento para estas medidas fueron ligeramente inferiores, pero considerando el número de recursos y la homogeneidad de las colecciones utilizadas en los experimentos mencionados por Lim y otros (2013), los resultados parecen igualmente prometedores.

Evaluación de la indexación realizada por humanos

Se pidió a seis expertos que extrajeran palabras clave de cinco documentos procedentes del mismo corpus utilizado para la evaluación automática. La tabla 4 contiene los resultados obtenidos por estos expertos, expresados de la misma manera que los resultados de la evaluación automática de indexación.

Tabla 4. Resultados obtenidos por los expertos humanos en términos de precisión, exhaustividad medida F.

Expert	Precision	Recall	F-measure
A	0.13	0.12	0.12
B	0.16	0.17	0.16
C	0.21	0.22	0.21
D	0.16	0.14	0.14
E	0.14	0.16	0.15
F	0.21	0.23	0.21
Human Average	0.168	0.173	0.171

Estos resultados muestran que cuando los expertos humanos tratan de hacer coincidir exactamente las palabras clave originales, sus resultados son ligeramente inferiores a los obtenidos por KEA. Aunque dos expertos (C y F) superaron a KEA, las desviaciones entre los mejores y peores resultados humanos no pueden ser consideradas significativas.

Evaluación manual por expertos de los resultados obtenidos

Siguiendo el procedimiento detallado en la sección 3.3.4, cada experto clasificó las palabras clave extraídas por sus colegas, las extraídas por KEA y las asignadas originalmente al documento. La tabla 5 presenta algunas de las calificaciones realizadas por el evaluador A. Cada evaluador informó sobre su evaluación en una forma similar a la que aparece en la tabla 5.

Tabla 5. Primeras filas de los resultados reportados por el evaluador A.

Evaluator A's Ratings					
Items	Completely wrong rep.=1	Not well rep.=2	Partially rep.=3	Well rep.=4	Perfectly rep.=5
Paper B1				X	
Paper B2				X	
Paper B3			X		
Paper B4				X	
Paper B5				X	
Paper C1		X			
(...)	(...)	(...)	(...)	(...)	(...)

La tabla 6 muestra los resultados que cada evaluador asignó a las palabras clave generadas por KEA. La Tabla 7 resume los resultados generales: la frecuencia de clasificaciones asignadas a las palabras clave seleccionadas por los evaluadores humanos, KEA y los indexadores originales.

Tabla 6. Resultados que cada evaluador humano asignó a las palabras clave extraídas por KEA

Results for KEA					
Evaluator	Paper 1	Paper 2	Paper 3	Paper 4	Paper 5
A	4	3	3	4	4
B	4	4	4	4	3
C	4	2	4	4	2
D	3	3	4	4	4
E	4	3	4	4	3

F	4	4	4	4	3
---	---	---	---	---	---

Tabla 7. Resultados globales de la evaluación manual.

Evaluator	Completely bad rep. (1)	Not well rep. (2)	Partially rep. (3)	Well rep. (4)	Perfectly rep. (5)
A	0	0	9	16	0
B	0	0	5	20	0
C	0	1	3	21	0
D	0	0	3	22	0
E	0	0	6	19	0
F	0	0	1	23	1
KEA	0	2	8	20	0
Original	0	1	5	24	0

La evaluación subjetiva y la tasa de acuerdo entre los sujetos participantes en el experimento fue evaluada utilizando el Coeficiente de Correlación Intraclass (ICC), que representa acuerdos entre dos o más evaluadores o métodos de evaluación sobre un conjunto de sujetos, manejando múltiples observadores y múltiples factores.

Fleiss y Cohen (1973) describieron la equivalencia de ICC con Kappa, también utilizada para evaluar la fiabilidad en experimentos anteriores sobre este tema y el coeficiente de correlación intraclass como medidas de fiabilidad. La ICC se ajusta a los efectos de la escala de mediciones, y al estar trabajando con una escala Likert de cinco puntos, se decidió emplear ICC para analizar la tasa acuerdo, como lo sugirió Norman (Norman, 2010). La tabla 8 muestra el análisis de la varianza de los resultados generales de la evaluación manual.

Tabla 8. Análisis de la varianza

Variance	df	Ssq	msq	F	p
B. Rows	4	2502.75	625.68	191.99	<0.0001
W. Rows	35	98.75	2.82		
B. Cols	7	7.5	1.07	0.32	0.8563
Residual err.	28	91.25	3.25		
Total	39	2601.5			

Con estos valores se calculó un ICC=0.965, lo que indica un nivel de acuerdo muy elevado, validando estadísticamente los resultados.

3.2.5 Limitaciones del experimento

Como se ha descrito anteriormente, se realizaron grandes esfuerzos para seleccionar una colección adecuada de documentos para el experimento de la mejor manera posible, lo cual incluyó varios procesos manuales, buscando este hecho. Para una mayor validación, el enfoque debería aplicarse en conjuntos más amplios de documentos, lo que requeriría la elaboración de herramientas capaces de elegir automáticamente documentos adecuados, descargarlos y comprobar su validez, es decir, evitar documentos directamente escaneados como imágenes, resúmenes y otros problemas similares.

Todos los sujetos participantes en este experimento eran expertos en el dominio. Debido a restricciones de tiempo y recursos, (la extracción de palabras clave utilizando un KOS es una tarea que requiere de una notable cantidad de tiempo, experiencia y conocimiento en el campo), tuvimos que limitar el alcance de este primer experimento.

3.2.6 Conclusiones del experimento

El objetivo principal de este experimento fue analizar la aplicabilidad práctica del servicio de indexación automática de KEA en un contexto real y práctico. Como se ha descrito, hasta la fecha, se han llevado a cabo un buen número de experimentos con KEA, pero la mayoría evalúa la eficacia tomando el conjunto de palabras clave originales asignadas manualmente, que pueden o no ser las "mejores"; como el patrón.

En algunos casos (El-Haj, Balkan, Barbalet, Bell, & Shepherdson, 2013) este enfoque es bastante problemático, y representa una clara limitación de los experimentos. Además, las muestras seleccionadas para experimentos anteriores han sido pequeños subconjuntos de colecciones, que contienen documentos sobre temas similares y / o desde el mismo período de tiempo. Este no es el caso de la colección AGRIS en VOA3R, de la cual se obtuvo la muestra en este estudio.

Considerando esto, se seleccionó un conjunto de documentos de la colección AGRIS en VOA3R para evaluar la efectividad de KEA según el criterio de precisión y exhaustividad, obteniendo resultados similares a los de experimentos anteriores, lo que proporciona pruebas adicionales de la eficacia de KEA en este tipo de tareas.

La comparación del desempeño de expertos humanos con KEA al intentar hacer coincidir exactamente las palabras clave asignadas originalmente mostró que KEA "superó" a los seres humanos, incluso empleando un corpus de entrenamiento de tamaño limitado (1200 documentos).

Este resultado no puede ser utilizado para concluir que KEA es mejor que los seres humanos, ya que la calidad de las palabras clave originales puede ser cuestionada, la diferencia de resultados no es altamente significativa y el número de evaluadores y artículos es limitado. Sin embargo, apunta claramente a que KEA proporciona resultados similares a los de los seres humanos.

La evaluación de concordancias exactas no proporciona información completa para fines de aplicabilidad práctica, ya que los términos semánticamente cercanos o con similitud morfológica también podrían ser útiles desde el punto de vista práctico. Por esta razón, se propuso un nuevo enfoque para validar las palabras clave, pidiendo a expertos humanos para evaluar palabras clave en cuanto a la calidad con la que clasifican un documento.

Los resultados obtenidos por KEA fueron similares a los obtenidos por humanos, lo que muestra su aplicabilidad práctica como un servicio de indexación automática para un repositorio que contiene recursos relacionados con el dominio cubierto por AGROVOC.

Dado que la elección de palabras clave para indexar recursos en una tarea compleja y que requiere de una notable cantidad de tiempo, especialmente cuando se utiliza un KOS de tamaño elevado como AGROVOC, KEA tendría aplicabilidad práctica también para apoyar el trabajo de los indexadores, por ejemplo, recomendando términos en el proceso de creación de metadatos o como una herramienta de aseguramiento de la calidad. Por último, se propuso un mecanismo de evaluación ciega, llenando un vacío en la investigación previa con KEA.

La ampliación del alcance del experimento en cuanto al número de documentos empleados, implicaría el desarrollo de un sistema capaz de buscar y descargar documentos de texto completo y luego transformarlos en un formato adecuado para KEA, lo cual incluiría un sistema de reconocimiento óptico de caracteres que permitiera trabajar con documentos escaneados.

En cuanto al análisis manual, la información podría perfeccionarse con detalles completos sobre el proceso de trabajo de los expertos para extraer palabras clave tales como el tiempo necesario, la competencia lingüística, la confianza en su trabajo, así como otros datos cualitativos; o más información sobre las características del documento: número de palabras, tipo de documento, año, calidad de la comunicación, etc.

Los esfuerzos futuros deben centrarse en las evaluaciones centradas en el usuario empleando el enfoque descrito en este documento, ya que la evaluación de la eficacia de KEA parece ser coherente en todos los experimentos realizados anteriormente.

3.3 Evaluación del grado de especificidad de dominio en los términos de un KOS

“Three Rings for the Elven-kings under the sky, / Seven for the Dwarf-lords in their halls of stone, / Nine for Mortal Men doomed to die, / One for the Dark Lord on his dark throne In the Land of Mordor where the Shadows lie. / One Ring to rule them all, One Ring to find them, / One Ring to bring them all and in the darkness bind them / In the Land of Mordor where the Shadows lie.”

(Extracto de “The Lord of the Rings: The Fellowship of the Ring”, de J. R. R. Tolkien , 1954)

3.3.1 Motivación del estudio

La creación de KOS suele venir motivada por la necesidad de organizar la información y limitar la complejidad de un dominio concreto, de manera que el KOS pueda ser aplicado para resolver problemas dentro de este dominio. Sin embargo, es frecuente que este tipo de KOS específicos de dominio, sobre todo aquellos de gran tamaño; puedan contener términos generalistas o que sean empleados en múltiples contextos, limitando la utilidad y la usabilidad del KOS.

El estudio que se describe a continuación profundiza en este hecho, proponiendo un método para evaluar la especificidad de dominio de los términos de los KOS, con el fin de poder proporcionar una respuesta a la tercera pregunta de investigación planteada en la presente tesis.

3.3.2 Introducción

Las terminologías extensas (Cabré, 2005) establecidas para un propósito o dominio en particular se organizan normalmente mediante la utilización de algún tipo de clasificación (Hjørland, 2001) a partir de la cual se generan diferentes sistemas de organización del conocimiento (Hodge, 2000).

Como ya se comentó en la revisión de la literatura, el uso de estos sistemas podría ser valioso para diferentes tareas relacionadas con la recuperación de la información y los propósitos de búsqueda (Aitchison, Gilchrist, & Bawden, 2000), tales como anotar e indexar los recursos utilizando vocabularios controlados (Park, y otros, 2013), aumentar las consultas de los usuarios mediante técnicas de expansión de consultas basadas en ontologías (Madhu, Govardhan, & Rajinikanth, 2011) o determinar el tema, en el sentido descrito por Hjørland (2001), de un documento (Paranjpe, 2009) (Gamon, Yano, X, Apacible, & Pantel, 2013).

Las grandes colecciones digitales generalmente utilizan terminologías grandes y KOS complejos, lo que complica las tareas mencionadas. Estas terminologías suelen incluir un gran número de términos generales que podrían aplicarse a otros dominios y podrían resultar irrelevantes o distraer a la hora de realizar las tareas anteriormente mencionadas.

Por ejemplo, el término "tecnología" se utiliza en el "Tesauro de Arte y Arquitectura" (AAT), pero el mismo término se puede encontrar en otros tesauros relacionados con dominios completamente diferentes, como el vocabulario controlado de "Medical Subjects Headings" (MeSH) y el vocabulario controlado agrícola AGROVOC.

En este trabajo se presenta un experimento que tiene como objetivo evaluar el grado de especificidad de dominio de grandes KOS, con el objetivo de extraer información relevante. En el futuro, este enfoque puede facilitar la búsqueda de información y las tareas de búsqueda en colecciones digitales específicas de dominio que empleen KOS compuestos por un elevado número de términos. Más específicamente, esta información podría ser útil para refinar las técnicas de expansión de consultas incluyendo los aspectos genéricos y específicos de un término como otro factor a considerar. También podría ser útil como una herramienta para reducir la complejidad de navegación de un KOS (Jules et al., 2013).

Los KOS de dominio específico contienen listas de términos específicos y genéricos que podrían ser útiles por sí mismos (Kim y Cavedon, 2011). Debe destacarse que el presente experimento se centra en las terminologías sobre las que se construyen KOS específicos de dominio, no con las relaciones particulares ni los esquemas de clasificación y categorización de los diferentes KOS.

Los experimentos se realizaron utilizando AGROVOC, el vocabulario controlado que abarca todas las áreas de interés para la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Como ya se indicó, el vocabulario de AGROVOC se utiliza para apoyar la recuperación de información y las tareas de búsqueda en la gran colección de AGRIS.

La evaluación propuesta en este documento se basa en un triple enfoque, que incluye análisis por expertos, cuantitativo y sistemático. Los resultados se contrastan de varias maneras para que los métodos complementen el análisis del estudio. El método de evaluación inicial del experimento siguió un enfoque de evaluación de expertos en el que se pidió a expertos en el campo que clasificaran los términos de AGROVOC indicando si cada término era específico del dominio de AGROVOC.

Antes de esto, se necesitaba un análisis en profundidad de AGROVOC para seleccionar una lista adecuada de términos para el experimento, que se explica en la sección "Selección de los materiales".

Una vez seleccionada una lista adecuada de términos, se pidió a los expertos que clasificaran esos términos y realizaran una evaluación para comparar los resultados y analizar cualquier discrepancia. Posteriormente, se propuso un criterio para medir el grado de especificidad de los términos en AGROVOC, teniendo en cuenta las particularidades de este KOS específico. Este criterio fue implementado para obtener un valor numérico final y definir el grado de especificidad de AGROVOC.

Desde el punto de vista de la aplicabilidad práctica, la evaluación de expertos humanos no siempre es posible debido al tiempo y los recursos necesarios, por lo que las evaluaciones automáticas se deberían considerar en esos casos.

La obtención de un conjunto de términos categorizados como resultado de la evaluación de los expertos humanos permitió hacer un primer acercamiento a la evaluación automática. Los resultados de los expertos humanos se tomaron como patrón, de manera análoga a como se realiza la evaluación de las técnicas automáticas de extracción de palabras clave (Lim, Wong, & Lim, 2013).

De esta manera, el conjunto de resultados establecido por los expertos humanos se contrastó con los términos en SUMO (ontología superior) y MILO (ontología de nivel medio) en un intento de examinar las posibilidades de utilizar estas ontologías como herramientas para determinar automáticamente el grado en que un término dado puede ser considerado específico del dominio.

Los resultados muestran la existencia de un alto número de términos genéricos y poco claros. La evaluación automática con SUMO y MILO evidencia que no hay una forma directa de evaluar el grado de especificidad usando sólo estas ontologías, ya que el número de términos concordantes no era suficiente. Sin embargo, esta evaluación proporciona información valiosa y una validación adicional de los resultados obtenidos por los expertos humanos, ya que el patrón de los resultados de la evaluación automática concuerda con las hipótesis iniciales.

3.3.3 Materiales y método

Este experimento se desarrolló en el contexto del dominio agrícola mediante el uso del vocabulario AGROVOC. AGROVOC es un vocabulario controlado que abarca los alimentos, la nutrición, la agricultura, la pesca, la silvicultura, el medio ambiente, la ganadería y todas las áreas de interés de la FAO. Fue desarrollado en los años ochenta para la base de datos del Sistema Internacional de Ciencia y Tecnología Agropecuaria (AGRIS) con el fin de estandarizar el proceso de indexación y hacer la búsqueda más sencilla y eficiente y guiar al usuario hacia los recursos más relevantes.

Para una mejor comprensión del enfoque de este experimento, es necesario destacar las siguientes consideraciones, que se han expuesto en el capítulo de la presente tesis dedicado al análisis de la literatura:

1. El paradigma del análisis de dominio desde la óptica de las ciencias de la información establece que la mejor manera de entender la información es estudiar los dominios de conocimiento como comunidades de pensamiento o discurso y la psicología personal: necesidades de información, conocimiento y criterios subjetivos (Hjørland & Albrechtsen, 1995).
2. Un "Dominio" debería reflejar una construcción social (Hjørland, 2001).
3. Para construir cualquier tipo de esquema de indexación para una comunidad en particular, uno puede estudiar cómo se hacen preguntas en ese ámbito (Mai 2004).
4. "Dominio" es un concepto evolutivo y abierto que se desarrolla a medida que el concepto se utiliza y, por tanto, también lo hacen los criterios subjetivos individuales de la persona "(Mai, 2005).

En consonancia con estos criterios, el experimento presentado evalúa el grado de especificidad del dominio de un conjunto de términos procedentes del KOS AGROVOC, como caso de una gran colección de términos específicos de dominio, basada en un estudio empírico (Hjørland, 2002) que toma en consideración los criterios subjetivos (Hjørland & Albrechtsen, 1995, mayo, 2005) de los sujetos participantes. Estos usuarios son expertos en el dominio, miembros activos de una comunidad (Hjørland, 2001; Mai, 2004) que usan y aplican términos relacionados con el dominio en la investigación y la práctica (Mai 2005).

Selección de los materiales

AGROVOC tiene 25 conceptos de nivel superior o raíz, es decir, descriptores sin ningún término más amplio. Dado esto, podemos argumentar que cada uno de estos conceptos constituye el punto de partida de una rama AGROVOC. El número de términos y la profundidad de cada rama varía significativamente. Se analizaron las 25 ramas, extrayendo todos sus términos y su profundidad para seleccionar las ramas más apropiadas para ser analizadas en el experimento.

La tabla 9 muestra el número de términos en cada rama iniciada por un concepto de nivel superior, proporcionando detalles sobre el número de términos más estrechos clasificados por profundidad. Se establecieron cuatro tamaños de ramas diferentes ("XL", "L", "M" y "S") según el número de términos de los conceptos. El idioma original en el que se realizó el experimento fue inglés, esto es, los términos empleados fueron los términos en inglés y la opinión recogida también se hizo en este idioma. Para evitar cualquier sesgo en la traducción de los datos, la información incluida en todas las tablas se muestra tal y como fue recogida.

Tabla 9. Los 25 conceptos raíz de AGROVOC: número de términos en cada rama, nivel de profundidad y tamaño asignado en función a los mismos.

	Terms	Size	D1	D2	D3	D4	D5	D6	D7	D8	D9	D10
Organisms	31,547	XL	1116	110	535	930	2203	5048	16976	1722	1823	1084
Substances	3372	L	58	229	658	692	882	604	215	34		
Entities	2476	L	90	301	431	345	397	572	223	103	14	
Phenomena	1716	L	27	173	344	405	494	218	53	2		
Activities	1415	L	88	490	365	183	147	112	26	4		
Features	1040	L	3	42	39	252	615	73	16			
Products	1040	L	17	140	409	357	106	11				
Methods	543	M	36	228	178	73	27	1				
Properties	458	M	27	192	175	49	14	1				
Objects	425	M	4	108	203	97	13					
Resources	416	M	8	17	29	89	50	214	9			
Subjects	397	M	5	33	80	87	86	80	21	5		
Systems	336	M	29	124	60	26	44	28	17	8		
Location	282	M	21	29	18	47	61	95	11			
Groups	266	M	41	97	70	48	7	3				
Measure	215	M	81	86	45	3						
State	163	M	46	57	41	16	3					
Stages	80	S	2	12	46	16	4					
Technology	73	S	20	12	20	16	3	2				
Processes	66	S	15	27	17	6	1					
Factors	59	S	11	14	25	7	2					
Time	24	S	6	15	2	1						
Site	17	S	10	7								
Events	17	S	5	7	4	1						
Strategies	8	S	3	5								

Las impresiones preliminares apuntaron a una particularidad con respecto a la jerarquía empleada en AGROVOC, ya que parece seguir un patrón con respecto a la especificidad del dominio de los términos.

Parecía que a medida que aumentaba la profundidad de los niveles de AGROVOC, los términos no sólo eran más específicos respecto a su término padre, sino también más "específicos del dominio". Por ejemplo, el término "Activities → Cleaning" es muy genérico, pero los términos más específicos bajo "Activities → Cleaning", como "Root Cleaning", "Grain Cleaning" o "Seed Cleaning", son no sólo más específicos respecto a "Activities" y "Cleaning", sino también más específicos dentro del dominio.

Debe destacarse que esta era sólo una circunstancia particular de AGROVOC, ya que, si bien es posible esperar un aumento en la especificidad de los términos a medida que se desciende en la jerarquía de KOS, no se puede afirmar que cada relación de tipo jerárquico esté directamente relacionada con la especificidad del dominio. La figura 10 representa los valores de profundidad para las cuatro categorías de tamaño de rama establecidas. Se puede observar que las curvas de las diferentes figuras representan una tendencia similar.

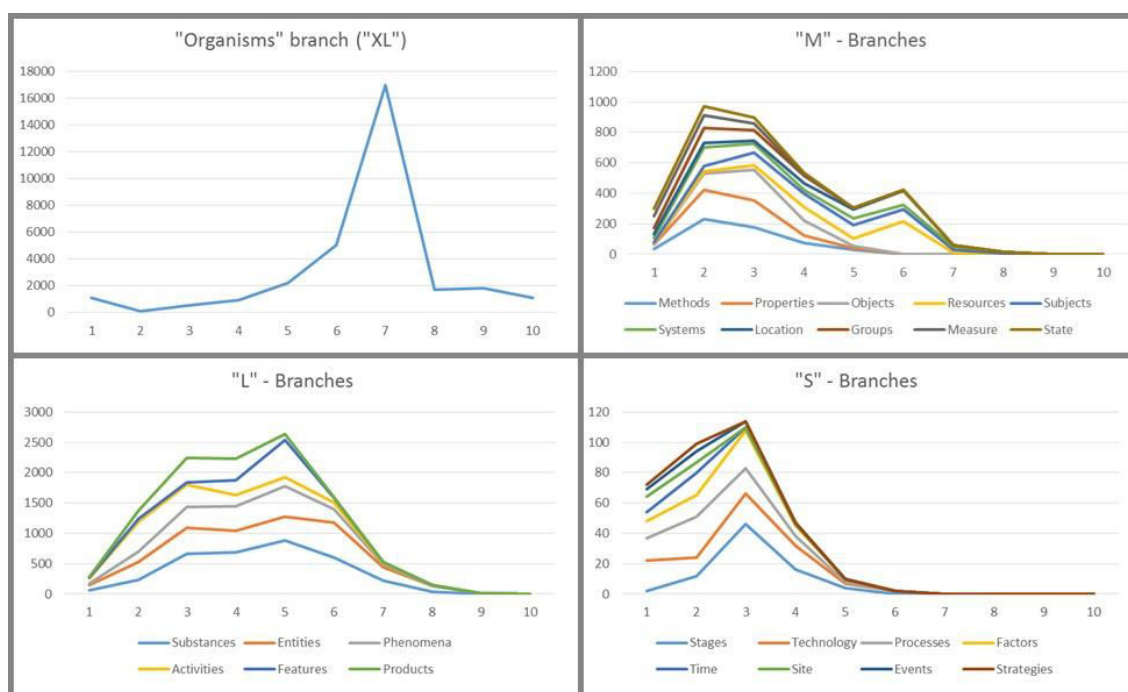


Figura 10. Número de términos clasificados por su profundidad para cada uno de los 4 grupos establecidos en función del tamaño asignado a la rama.

Además del análisis estadístico, también se analizó cada rama desde un punto de vista cualitativo para descubrir detalles que hacen que algunas ramas sean más adecuadas para el experimento que otras. Por ejemplo, las ramas tales como las encabezadas por los términos "organismos" o "sustancias" eran "altamente específicas" en el sentido que contenían nombres científicos de diferentes especies y sustancias relacionadas con los dominios de AGROVOC, por lo que no podrían proporcionar información relevante para el experimento.

Estas consideraciones se tuvieron en cuenta para hacer una selección adecuada de términos para el experimento. Se consideraron los conceptos de nivel superior con el mayor número de términos finales dentro de cada una de las ramas de diferente tamaño, que parecieran contener un número significativo de términos específicos y genéricos, y que contuvieran términos al menos hasta seis niveles de profundidad.

El objetivo era obtener una muestra total de alrededor de 2000 términos, teniendo en cuenta que la evaluación manual de un número de términos supondría un tiempo y esfuerzo considerable por parte de los expertos que participan voluntariamente en el experimento. Por lo tanto, el criterio de selección final para las ramas más adecuadas se hizo sobre la base del objetivo de obtener una muestra con un número de términos cercano a 2000.

La evaluación llevó a seleccionar las ramas de “Technology” (73 términos), “Properties” (458 términos) y “Activities” (1415 términos) como las ramas más adecuadas, obteniendo una muestra total de exactamente 1946 términos. La tabla 10 resume las conclusiones de la evaluación interna para la selección de los términos principales.

Tabla 10. Conclusiones de la evaluación interna para la selección de los términos.

Top Term	# Terms	Max depth	Qualitative comments
Organisms	31547 - too many		
Substances	3372 - too many		
Entities	2476 - too many		
Phenomena	1716 - too many		
Activities	1415	8	Suitable
Features	1040	7	Mostly contains names of rivers, lakes and other geographical features.
Products	1040	6	A list of products. Very specific.
Methods	543	6	Suitable
Properties	458	6	Suitable
Objects	425	5	
Resources	416	7	A list of natural resources - including specific names of rivers, gulfs, etc.
Subjects	397	8	Most of these are generic subjects, not related to AGROVOC domains.
Systems	336	8	Suitable
Location	282	7	Mostly contains names of different locations.
Groups	266	6	Suitable
Measure	215	4	
State	163	5	
Stages	80	5	
Technology	73	6	Suitable
Processes	66		
Factors	59 - too few		
Time	24 - too few		
Site	17 - too few		
Events	17 - too few		
Strategies	8 - too few		

Cabe destacar que el experimento no persigue la evaluación del grado de especificidad de dominio de todo AGROVOC, lo que requeriría el análisis de todo el KOS, sino proponer una metodología para evaluar el grado de especificidad de dominio de cualquier terminología. Por lo tanto, también podrían haberse considerado adecuadas otras ramas u otras clases de selecciones de términos procedentes de terminologías diferentes.

Recolección de datos

Los datos se recopilaron mediante el uso de un cuestionario que contenía todos los términos. Se solicitó a los expertos que participaron en el experimento que clasificaran cada término como "específico" o "genérico". En los casos en los que no estuvieran seguros acerca de la especificidad del término se alentó a los expertos a calificar si el término era más probable que fuera genérico o específico e indicar la razón, con el fin de obtener información cualitativa sobre estos casos. Para calificar los casos poco claros, los participantes usaron una escala de 1 a 9, significando 1 "más cercano a ser genérico" y 9 "más cercano a ser específico", 10 fue considerado como "completamente seguro de ser específico" (S) y 0 "completamente seguro de ser genérico" (G). La tabla 11 muestra un ejemplo abreviado del cuestionario.

Tabla 11. Ejemplo abreviado del cuestionario con el que se obtuvieron los datos, el fragmento mostrado corresponde a un caso real.

Term	S/G	Not clear rate	Observations
Recursive programming	G		<i>"Seems to be quite computer-specific, but not agriculture-specific."</i>
Diagnosis	G		
UNCED		5	<i>"An organization related to agriculture? This could mean anything..."</i>
Seed cleaning	S		

Con el fin de evitar cualquier tipo de influencia en las decisiones de los expertos, se les proporcionó una hoja de cálculo que contenía sólo una lista de términos completamente descontextualizada, en el sentido que esta no contenía información sobre la jerarquía o las ramas que parten de los términos principales, tal como se muestra en la tabla 12.

De este modo se obtuvo un conjunto de datos que contenía los valores asignados a cada término por 16 expertos. Los comentarios en el campo "Observaciones" fueron muy útiles para discernir las razones de las discrepancias de los expertos al calificar los términos. Algunos de los expertos hicieron comentarios sobre diferentes términos, mientras que otros proporcionaron comentarios generales.

Después de un primer análisis, se decidió contactar directamente a algunos expertos para:

- Pedir más información para arrojar luz sobre la información dada en el campo "Observaciones", donde las discrepancias permanecen poco claras.
- Confirmar nuestras suposiciones.

Las respuestas se resumen en la sección "Resultados", que también proporciona detalles sobre las discrepancias de los expertos.

Contraste de los resultados con ontologías de propósito general

La última parte del experimento contrastó el conjunto de datos obtenidos a partir de los resultados de la evaluación de expertos con la ontología superior SUMO (Niles & Pease, 2001) y la ontología de nivel medio de propósito general MILO (Niles & Terry, 2004), la cual pretende ser un nexo entre el contenido abstracto de SUMO y la especificidad de las diversas ontologías de dominio.

Cabe señalar que no se comparó directamente AGROVOC (tesauro) con SUMO o MILO (ontologías). De esta parte del experimento se esperaba obtener una primera impresión con respecto a la utilidad de emplear SUMO y MILO para apoyar la evaluación automática de la especificidad de términos de dominio. Para ello, se utilizó el conjunto de datos resultante de la evaluación de expertos humanos sobre AGROVOC como patrón clave.

3.3.4 Discusión de los resultados

Resultados de especificidad según la evaluación de los expertos

Todos los resultados fueron recogidos en una sola tabla, asignando los valores numéricos antes mencionados para calcular el "grado de especificidad" de cada término. A continuación, se calculó la media para cada término y se utilizó este resultado para clasificar cada término como específico o no, de acuerdo con la siguiente escala:

- [0-3]: Genérico / No específico del dominio
- [3-7]: Incierto / No claro
- (7-10]: Específico del dominio

Se consideró innecesario el uso de criterios de calificación más complejos o de cualquier otro instrumento estadístico debido a que una media dentro de los valores [3-7] representaría uno de los siguientes casos:

- A) Número similar de ocurrencias de valores altos y bajos. Este caso representa una tasa muy baja de acuerdo, dando lugar a un resultado "no claro".
- B) Todos los valores están en el rango [3-7], lo que significa que los expertos no pueden categorizar claramente los términos como específicos o no. Este caso nos lleva también al resultado "No está claro".
- C) Una combinación de (a) y (b).

La tabla 12 muestra los valores asignados por expertos, la media calculada y su categorización de especificidad de dominio para algunos de los términos analizados.

Tabla 12. Valores asignados por expertos, media calculada y categorización de especificidad de dominio para algunos de los términos analizados

TERM/EXPERT	1	2	3...14	15	16	MEAN	Domain Specific
Accounting	0	0		0	0	1,25	No
Auditing	0	0		0	0	0,625	No
Depreciation	0	0		0	10	1,5625	No
Seed certification	10	10		10	10	10	Yes
Choice of species	10	10		10	0	9,375	Yes
Dust extraction	10	10		10	10	6,875	Not clear
Grain cleaning	10	10		10	10	10	Yes

La figura 11 muestra los "resultados de especificidad" para las ramas analizadas, comparando el porcentaje de términos claros y de este el porcentaje de términos específicos.

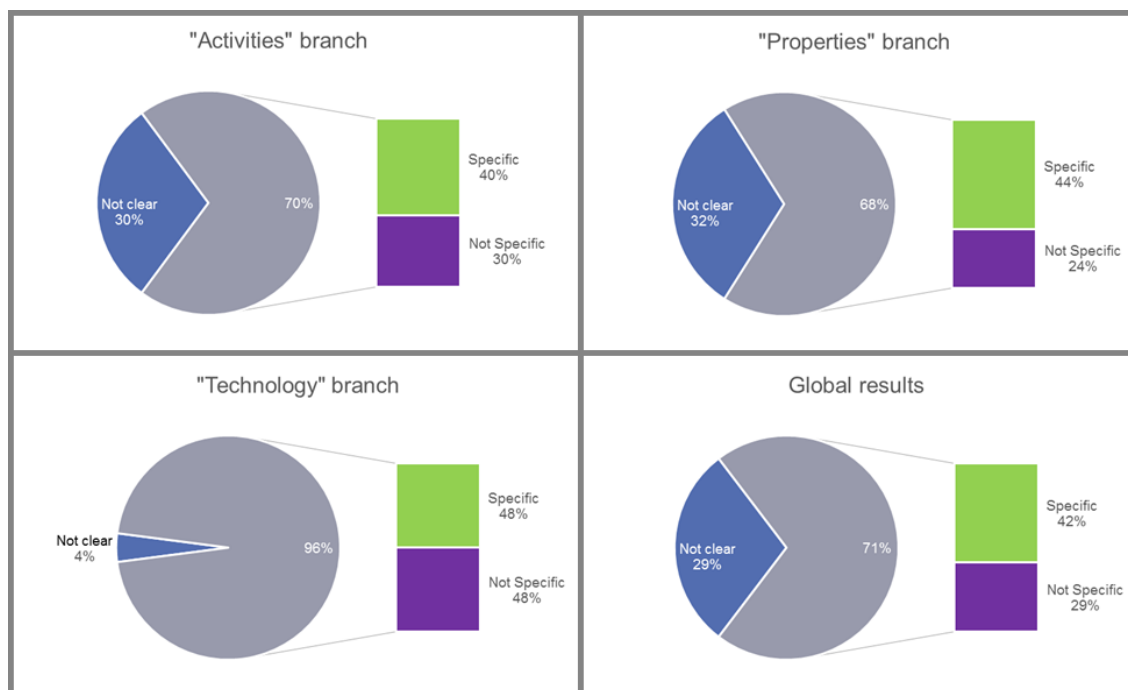


Figura 11. Resultados de especificidad de los términos en porcentaje, por ramas.

Como se comentó anteriormente, de acuerdo con la relación jerárquica empleada, sería lógico asumir que a medida que aumenta el nivel de profundidad, los términos son más específicos dentro del dominio; la figura 12 muestra los resultados en cuanto a este asunto.

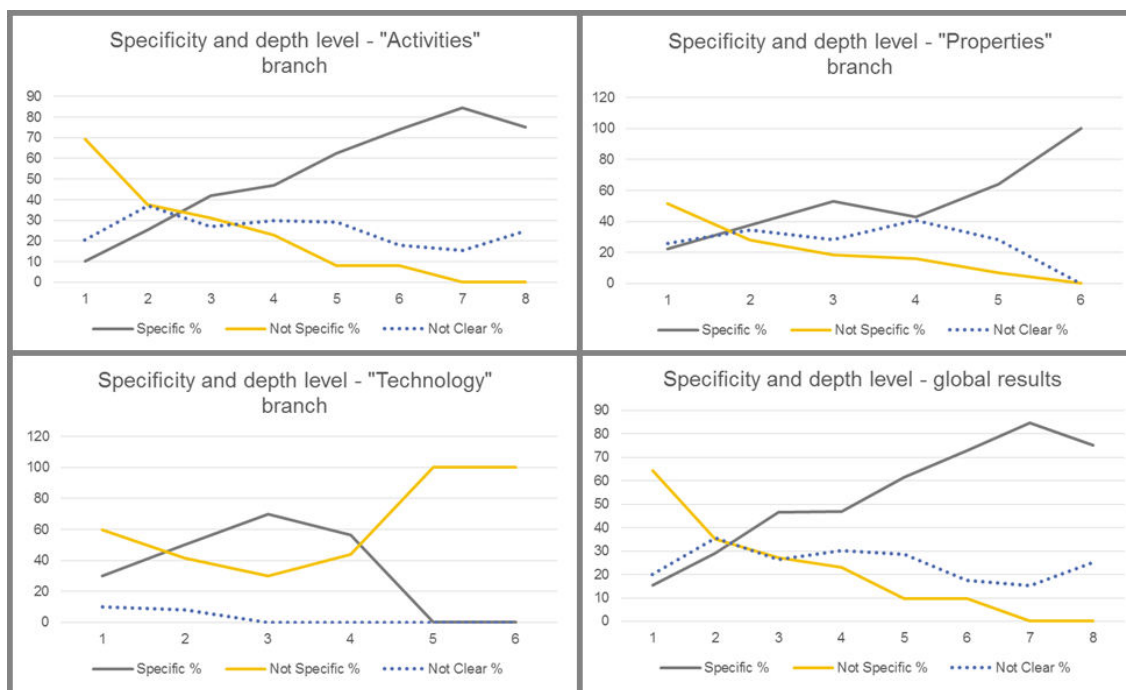


Figura 12. Resultados de especificidad de los términos en porcentaje (ordenadas) en relación al nivel de profundidad (abscisas).

Los resultados de las ramas "Activities" y "Properties" parecen validar la suposición original, pero no es el caso de la rama "Technology", ya que todos los términos con una profundidad de "5" y "6" son genéricos. El análisis de este hecho debe tener en cuenta que la rama "Technology" es la más pequeña, contiene sólo 73 términos y la totalidad de los términos de mayor profundidad son totalmente genéricos. La figura 13 ilustra esta situación.

A pesar de la discrepancia en la rama "Tecnología", los resultados globales, que incluyen los términos de las tres ramas, confirman la suposición original, como se muestra en la figura 12. Con esto, podemos comprobar que al observar otras ramas en el grupo "XL" O "L", o mirando todos los términos hasta cierto nivel, podrían generar resultados más consistentes que los obtenidos con el análisis de las ramas "S".

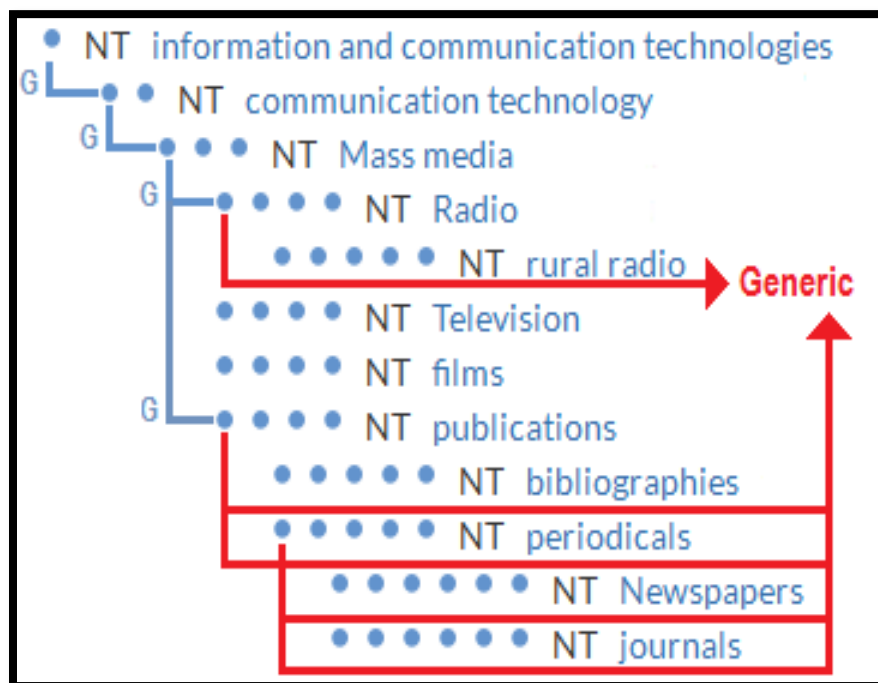


Figura 13. Todos los términos de este conjunto, proveniente de la rama “Technology” son genéricos, los términos con profundidad “5” y “6” están subrayados en rojo.

Casos inciertos

Con el fin de obtener mayor información sobre el razonamiento de los expertos detrás de la asignación del "grado de especificidad" de un término se realizó un análisis cualitativo. Para ello, los casos poco claros y los comentarios de los expertos se analizaron y clasificaron primero y luego se contactó con algunos de los participantes directamente para obtener más aclaraciones. La tabla 13 enumera los casos inciertos y describe sus posibles causas, según el análisis realizado:

Tabla 13. Casos inciertos y motivación detectada para cada uno de ellos tras el análisis cualitativo.

Caso		Motivación
a	Acronyms such as “UNCED.”	<i>They could refer to other institutions, technologies, processes, materials, etc.</i>
b	Specific terms but from other areas, such as computer science/IT. For instance, “recursive programming.”	<i>Some experts got confused; they were told to mark with “G” not only the generic terms, but those not specific to the AGROVOC domains, as depicted in table 3.</i>
c	Specific terms but from other areas, such as economics. For instance, “environmental accounting.”	<i>Same as case “b”</i>
d	Specific terms but from other areas, such as learning / education. For instance, “vocational education.”	<i>Same as case “b”</i>
e	Monitoring and measuring processes such as “gravimetry”, or terms related to weight.	<i>They could refer to other areas, even though they are widely used in the domains of agriculture and animal husbandry.</i>
f	Chemical and physical processes such as “electrodialysis” or “decanting.”	<i>Similar to case “e”</i>
g	Medical terms such as “pregnancy diagnosis” and several pertaining to blood composition or toxicity.	<i>They are related to animal husbandry in an AGROVOC context, but they are also specific to the human healthcare domain.</i>
h	Terms such as “cultivation” or “forcing.”	<i>Have different acceptations which have specific meaning(s) in other domain(s).</i>
i	Names of common fruits and vegetables. For instance, “apple.”	<i>There was no agreement about whether these terms should be considered specific (since they are clearly related to the agricultural and food domains) or generic (since they are widely used terms).</i>
j	Related to fire. For instance, “fire control.”	<i>These terms are mostly related to forestry in AGROVOC context, but some experts consider that they could be used in other domains.</i>
k	Work-related terms. For instance, “stacking”, “pneumatic handling”, or “unloading.”	<i>They are specific terms related to handling and working. They are widely used in agricultural and forestry domains, but they could be used in many other areas.</i>
l	Shipping and handling-related terms, such as “packaging” or “wrapping.”	<i>Mostly related to food handling in the AGROVOC context but they could refer to other areas.</i>
m	Ways to prepare food, such as “smoldering” or “candyng.”	<i>Some experts consider these terms to be “too common” and more related to cooking than to food and nutrition domains.</i>

Cálculo del grado de especificidad de dominio de las ramas de AGROVOC analizadas

Considerando lo expuesto anteriormente sobre la relación entre el número de términos específicos y la "profundidad" de los términos, se propuso un método para calcular el grado de especificidad de dominio para cada rama de AGROVOC analizada, calculando diferentes pesos para los términos según su profundidad.

Cabe señalar que este enfoque fue elegido debido a las particularidades de la jerarquía AGROVOC y, por lo tanto, no tiene porqué ser directamente aplicado a otros KOS. Para calcular adecuadamente el grado de especificidad siguiendo nuestro criterio en otros casos, el peso debe ser calculado tomando en cuenta otras particularidades posibles del KOS en cuestión o ponderando todos los términos de manera igual.

En primer lugar, se calcularon distintos pesos (W) para los términos de acuerdo a su nivel de profundidad (D). En el nivel más profundo (D_M) se consideró tener un peso = 1. Así, el peso del término con la profundidad " x " (D_X) será:

$$W_x = \frac{D_x}{D_M}$$

De acuerdo a esto, la puntuación máxima de especificidad (S_M - "*Maximum Specificity Score*") será:

$$S_M = \sum_{x=1}^{D_M} T_x * W_x$$

Donde T_x es el número de términos con profundidad = x .

Siguiendo el mismo criterio se calcularon la puntuación de especificidad real (S_r - "*Real Specificity Score*"), la puntuación para los casos inciertos (S_u - "*Unclear Specificity Score*") y la puntuación para los casos claros (S_c - "*Clear Specificity Score*"), de la siguiente manera:

$$S_r = \sum_{x=1}^{D_M} T_{sx} * W_x \quad S_u = \sum_{x=1}^{D_M} T_{ux} * W_x \quad S_c = S_M - S_u$$

Donde T_{sx} es el número de términos considerados como específicos con profundidad = x y T_{ux} es el número de términos considerados como inciertos con profundidad = x.

Finalmente se realizó una analogía con el criterio de precisión / exhaustividad propuesto por Van Rijsbergen (1979) con el fin de obtener el resultado de especificidad global. El número de términos específicos se expresa como una proporción de todos los términos (*precisión "P"*) y el número de los términos claramente clasificados (*acuerdo "A"*) para cada rama, analizada de forma individual, de la siguiente manera:

$$P = \frac{S_r}{S_M} \qquad A = \frac{S_r}{S_c}$$

El resultado final o grado de especificidad (S - "*Specificity Degree*") es calculado como una combinación de ambos:

$$S = \frac{2PA}{P + A}$$

La tabla 14 muestra los resultados calculados siguiendo este criterio.

Tabla 14. Resultados de especificidad obtenidos tras la aplicación del criterio propuesto.

Rama	S_M	S_r	S_u	S_c	P	A	S
<i>Technology</i>	32,5	16	0,66	31,83	0,492	0,502	0,497
<i>Properties</i>	201,33	94	64,83	136,5	0,466	0,688	0,556
<i>Activities</i>	549,5	270,37	154,87	384,12	0,492	0,703	0,579

Las conclusiones obtenidas deben tener en cuenta que ($P, A, S \leq 1$), o en otras palabras, que el valor máximo que P, A , y S pueden tomar es 1. Si los valores de P y A difieren ampliamente, esto será debido al número de casos poco claros o inciertos. Este grado de incertidumbre (I), puede ser calculado del siguiente modo:

$$I = \frac{S_u}{S_M}$$

Comparación de los resultados con las ontologías SUMO y MILO

El experimento con expertos humanos mostró que evaluar la especificidad de una lista de términos es una tarea que requiere mucho tiempo y es más complejo de lo que pareciera a primera vista. Por ello se indagó sobre la posibilidad de emplear mecanismos automatizados para llevar a cabo la tarea. Este primer acercamiento consistió en contrastar los resultados obtenidos por los expertos humanos con las ontologías SUMO y MILO. La hipótesis inicial fue que, como SUMO y MILO no son ontologías específicas, los términos considerados "específicos" por expertos humanos no deberían existir en las mismas. La tabla 15 muestra los resultados específicos, no específicos e inciertos contrastados con estas ontologías.

Tabla 15. Resultados de contrastar el set de términos elaborado en base a la evaluación de los expertos humanos con los términos existentes en las ontologías SUMO y MILO.

	Not Specific	Not Clear	Specific
	<i>HUMAN EXPERTS (Gold Standard)</i>		
Terms found	567	570	809
	<i>SUMO</i>		
Terms found	65(11,463%)	13(2,28%)	0(0%)
	<i>MILO</i>		
Terms found	108(19,047%)	36(6,315%)	6(0,741%)

La cantidad de términos encontrados en SUMO y MILO demuestra claramente que no es posible utilizar directamente SUMO o / y MILO como un mecanismo automático para evaluar el grado de especificidad de un término. Sin embargo, los resultados validan los supuestos iniciales en el sentido que el número de términos no específicos encontrados fue significativamente mayor que el de términos inciertos y mucho más alto en comparación con el número de términos específicos encontrados. Esto también se refleja en el hecho de que los términos encontrados en MILO fueron más que en SUMO y con una distribución similar, como se muestra en la figura 14.

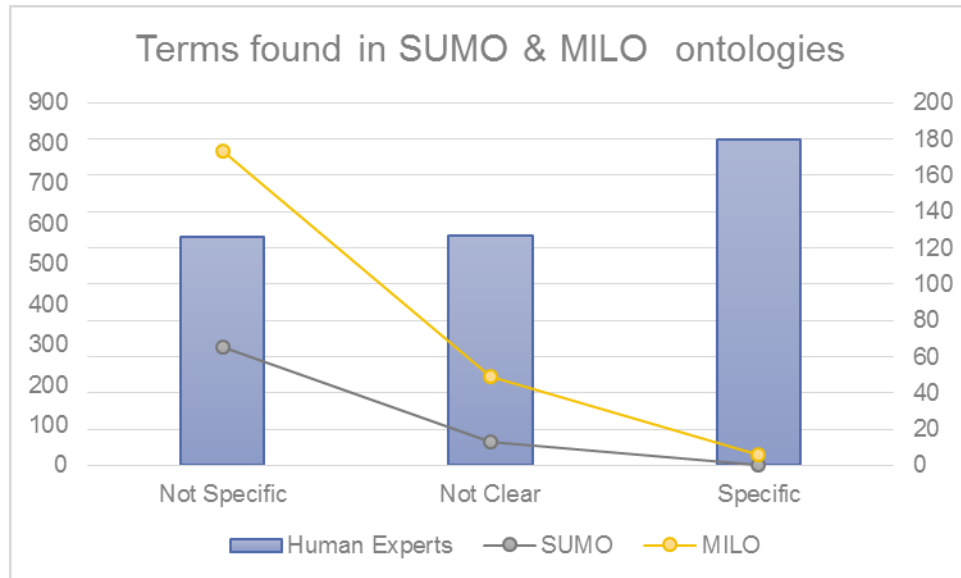


Figura 14. Términos coincidentes con las ontologías SUMO y MILO (escala de la derecha - 0,200) con los extraídos en la evaluación realizada por expertos humanos (escala de la izquierda - 0,900)

3.3.5 Limitaciones del experimento

Cabe señalar que, a pesar de que se pidió a los expertos que clasificaran los términos en pequeños conjuntos y se diera más de un mes para este proceso, no podemos descartar la posibilidad de error humano o malentendido de algunos términos. De hecho, en el proceso de la entrevista de seguimiento, algunos expertos no pudieron aclarar algunas de sus decisiones, aunque el número de estos casos parecía ser insignificante según nuestro análisis. Ciertamente, la recolección de datos de más expertos sería útil para aclarar algunos casos, pero desafortunadamente, no fue posible aumentar el alcance del experimento.

3.3.6 Conclusiones del experimento

Evaluar manualmente el grado de especificidad de terminologías es una tarea que requiere de una elevada cantidad de tiempo y es más complejo de lo esperado a primera vista. El experimento ha demostrado que los términos están sujetos a diferentes interpretaciones y opiniones subjetivas que llevan a los expertos a desacuerdos notorios o simplemente no les permiten ser claramente clasificados como de dominio específico o genéricos.

Esta clasificación podría utilizarse para facilitar la recuperación de información y las tareas de búsqueda en colecciones digitales específicas de dominio que empleen terminologías de gran tamaño, tomando en consideración la especificidad de un término como otro factor a considerar en la aplicación de las distintas técnicas existentes. Esta información también podría ser empleada a la hora de tratar de inferir el tema o dominio de un documento.

Se propuso un enfoque para evaluar manualmente la especificidad de una terminología de gran tamaño, el cual se probó con diferentes ramas de AGROVOC, proporcionando gran detalle sobre todos los procesos utilizados y las decisiones tomadas. Además, se ha propuesto un método para calcular el grado de especificidad, expresando el resultado final de este cálculo en una escala de 0 a 1. Este cálculo considera la "profundidad" de cada término asignando diferentes pesos, tomando en cuenta la relación jerarquía empleada por el KOS.

Hubo un número considerable de casos no claros al clasificar los términos, por lo que se realizó un análisis cualitativo para agrupar estos casos y describir las posibles razones de la ambigüedad.

El conjunto de resultados procedentes de la evaluación experta se contrastó con las ontologías SUMO y MILO, suponiendo que los términos considerados "específicos" por expertos humanos no deberían existir en estas ontologías no específicas. Este enfoque constituyó un primer acercamiento para distinguir las posibilidades de automatizar la clasificación de términos de dominio específico y nos proporcionó resultados valiosos puesto que, a pesar de que el número de términos concordantes era demasiado bajo para usar directamente SUMO o MILO como un mecanismo automático para evaluar el grado de especificidad de un término, el patrón de los resultados siguió los supuestos iniciales y proporcionó validación adicional a los resultados obtenidos por los expertos humanos.

Los siguientes pasos deberían centrarse en la búsqueda de mecanismos automáticos para evaluar el grado de especificidad de dominio. Un acercamiento para encontrar este método de evaluación automático pasaría por el uso de ontologías de nivel medio y alto, como SUMO y MILO, combinado con bases de datos léxicas como WordNet y otras fuentes de información como DBpedia. Además, el criterio propuesto para calcular el valor del grado de especificidad debe someterse a nuevas pruebas a fin de aclarar en qué medida podría utilizarse en otros contextos y con otros KOS.

3.4 Evaluación de la distancia de especialización entre conceptos

“Tuve un hogar y una daga. Conozco la esgrima, la gramática, las cuatro reglas y algo de latín. Sé escribir con buena letra, a veces leo libros y he visto mundo... ¿Qué más puede pedir el huérfano de un soldado de Flandes?”

(Extracto de “El puente de los Asesinos”, colección Capitán Alatríste, A. Pérez-Reverte, 2011)

3.4.1 Motivación del estudio

En su investigación, Sicilia y otros (2003) indican que el entendimiento común de las relaciones del tipo “Generalización / Especificación” asume que la relación entre un concepto y cualquiera de sus conceptos relacionados siempre es igualmente fuerte y también en todos los niveles de la jerarquía; lo cual supone una simplificación excesiva de las relaciones del mundo real, la cual podría ser limitada asignando valores a las relaciones que identifiquen la distancia de especialización entre los conceptos.

De hecho, en la evaluación plasmada en la sección anterior, se analiza cómo evoluciona la especificidad de dominio a medida que se desciende en el árbol jerárquico de dominio, el cual se establece en base a una relación del tipo “Generalización / Especificación”, observándose la existencia de términos hermanos que presentaban un nivel de especificidad llamativamente distinto respecto al término padre.

Esta apreciación en cuanto a la distancia de especialización entre conceptos observada en AGROVOC respalda la motivación de realizar un estudio para analizar los beneficios del empleo de esta distancia a través de una evaluación con usuarios (Martín-Moncunill, García-Barriocanal, Sicilia, & Stracke, en prensa), pudiendo aprovechar parte del análisis realizado en el experimento anterior. Concretamente, el trabajo en cuanto al análisis de las ramas para la selección de los términos más adecuados para la realización del experimento anterior resulta de gran utilidad para el experimento descrito en esta sección.

Para poder proporcionar una respuesta a la cuarta pregunta de investigación, relativa a la posibilidad de mejorar la usabilidad de un KOS asignando un grado de distancia de especialización entre conceptos, es necesario el desarrollo de un método para asignar esta distancia a un conjunto de términos de AGROVOC, puesto que no existe ningún KOS que incluya dicha información. Una vez obtenido un KOS con esta información, es posible implementar interfaces que incluyan esta información y evaluarlas siguiendo un test A / B, comparándolas con casos análogos que no hagan uso de la distancia de especialización entre conceptos, comprobando así si los usuarios encuentran utilidad en emplear esta información.

3.4.2 Introducción

Como se indicó anteriormente, la construcción de KOS como tesauros u ontologías, es una tarea compleja que normalmente comienza generando un corpus de términos (Kim y Cavedon, 2011). Una vez establecido este corpus, los términos se pueden estructurar estableciendo relaciones entre ellos. En este sentido, un KOS es una forma de representación del conocimiento que tiene como objetivo organizar la terminología para un propósito particular.

La relación de generalización / especialización ("gen-spec") es una relación entre clasificadores (términos o conceptos) que implica una relación taxonómica y su subsiguiente semántica hereditaria. Según Aitchison y otros (2000), los tesauros emplean este tipo de jerarquía, denominada en inglés como "*broader / narrower*" (ISO 25964, 2013) y de manera similar, las ontologías usan la relación conocida como "is-a" para representar este tipo de jerarquía.

El entendimiento común de este tipo de relaciones las considera como "todo o nada", asumiendo que la relación siempre es igualmente fuerte entre un clasificador y cualquiera de sus clasificadores relacionados mediante una relación "gen-spec" y también en todos los niveles de la jerarquía. Esto ha sido considerado por autores como Cohen y Murphy (2003), Sicilia y otros (2003), Cross (2004) o Hu y otros (2007) una simplificación excesiva del relato psicológico de las relaciones del mundo real.

La asignación de un grado de distancia relativa para representar el nivel de similitud entre los pares de clasificadores relacionados podría ser valiosa para fines de búsqueda (Sicilia, García, Aedo, & Díaz, 2003) y se podría aplicar para la representación visual de KOS. Respecto a la recuperación de la información, esta distancia puede utilizarse para establecer pesos que se tengan en cuenta para una mejor toma de decisiones sobre la sugerencia de términos de búsqueda o resultados relacionados. En cuanto a interfaces visuales, se podría utilizar para decidir la representación de los diferentes clasificadores y su posición en la pantalla.

En el experimento presentado en esta sección se evalúa la "distancia de especialización entre conceptos" descrita por Sicilia et al., empleando AGROVOC. El primer paso consistió en un análisis de AGROVOC para encontrar una muestra adecuada para el experimento, donde se aprovechó el análisis realizado en el experimento descrito en la sección anterior. A continuación, se pidió a un grupo de 21 personas asignar valores a las diferentes relaciones, para obtener un conjunto de datos consistente de distancias de especialización entre conceptos desde una perspectiva de usuario final y asignar de este modo pesos definitivos a las relaciones.

Con el objetivo de medir el impacto desde la perspectiva del usuario, se construyeron dos conjuntos de representaciones de las relaciones entre términos, uno según el concepto calculado de pesos de especialización y el otro siguiendo el orden original del tesauro (alfabético). Finalmente, se pidió a cuarenta personas que eligieran entre las versiones en un

experimento tipo A / B y se realizaron entrevistas cortas después de la prueba para investigar sus decisiones.

Los resultados muestran que el uso de esta información podría ser una valiosa herramienta para la búsqueda y recuperación de información y para la representación visual de KOS. Además, la metodología seguida en el estudio resultó ser útil para detectar incoherencias en el KOS y podría así ser utilizada para el control de calidad y la optimización de las relaciones jerárquicas.

3.4.3 Materiales y método

Los conceptos existentes en AGROVOC se organizan en una jerarquía cuyas relaciones quedan definidas del siguiente modo:

- NT (*"narrower term"*): Si X es un NT de Y, entonces X es más específico en algún sentido que Y. Por ejemplo, *"Food preservation"* (preservación de alimentos) NT *"Brining"* (salmuera).
- BT (*"broader term"*): Si Y es un BT de X, entonces X es más amplio que Y. Por ejemplo *"Brining"* BT *"Food preservation"*. BT es la inversa de NT.
- RT (*"related term"*): Esta relación expresa cualquier tipo de relaciones asociativas entre dos términos que no sean jerárquicos. Es el predeterminado para todas las relaciones, por lo que es muy ambiguo.
- UF (*"used for"*): Es una relación de equivalencia que indica que el término X es USADO PARA el término Y, con propósitos de indexación.

En este sentido, las relaciones NT y BT son relaciones jerárquicas que siguen el concepto de "gen-spec", haciendo de AGROVOC un KOS adecuado para el experimento.

La primera etapa del experimento consistió en analizar AGROVOC para encontrar una muestra adecuada de términos. Concretamente, se buscó un conjunto de términos de uso común, con el objetivo de asegurar que la falta de experiencia en el dominio no sesgara los resultados y se pudiera acceder fácilmente a una muestra más amplia para el experimento.

Siguiendo este enfoque, se asignaron valores de distancia de especialización entre los conceptos seleccionados en base a la evaluación realizada por 21 participantes que asignaron subjetivamente valores de distancia de especialización usando una escala de 5 puntos. El análisis de estos resultados condujo a la elaboración de un conjunto de datos que contenía los términos y las distancias de especialización adecuados para la evaluación de la interacción con usuarios empleando esta información en interfaces. Esta evaluación siguió un enfoque de prueba A / B, donde los sujetos tenían que elegir entre los casos que empleaban la información sobre distancia de especialización entre conceptos y los que no.

Selección de los materiales

El análisis de las secciones de AGROVOC recogido en el experimento anterior mostró que no solamente el número de términos y niveles de jerarquía varía significativamente, sino también su especificidad de dominio y complejidad para ser entendido por sujetos no relacionados con el dominio.

Se observó que las ramas encabezadas por términos como "organismos" o "sustancias" eran "altamente específicas" por incluir en su práctica totalidad nombres de diferentes especies y sustancias relacionadas con los dominios AGROVOC (ver tabla 9). Este tipo de ramas no eran adecuadas para el experimento, ya que se requeriría un alto dominio de conocimiento para el proceso de evaluación, lo cual se alejaba del enfoque del experimento actual.

Como se expuso anteriormente, el propósito del experimento no era centrarse sólo en un tipo de perfil de usuario (es decir, experto en el dominio agrícola), ya que esto podría limitar los resultados del experimento. Además, debe tenerse en cuenta que el número de relaciones a evaluar debe ser adecuado para las personas que participan voluntariamente en el experimento, en el sentido de que podrían sesgar su evaluación por cansancio, tedio o simplemente por sentirse abrumado por el número de términos, relaciones o niveles jerárquicos.

Por la misma razón, considerando que los participantes en el experimento no eran expertos en organización del conocimiento, se empleó una escala de 5 puntos para evaluar su apreciación subjetiva de la distancia de especialización entre conceptos.

El análisis, basado en las conclusiones obtenidas en el experimento recogido en la sección anterior, condujo a encontrar una subsección de términos comúnmente utilizados en la rama "Products", anidados bajo el término "Foods".

Debe observarse que podría haber otras selecciones de términos que podrían ser adecuadas para el experimento, pero elegimos estas por razones de familiaridad y simplicidad y teniendo en cuenta que el número de términos anidados bajo "alimento" no abrumaría a los sujetos participando en los experimentos.

La tabla 16 muestra los términos en el segundo nivel de la jerarquía (es decir, anidados bajo el término "Alimento"); puede apreciarse fácilmente que la mayoría de estos términos son comunes, evitando problemas de comprensión que podrían surgir si se usan otras ramas que contienen un elevado número de términos específicos del dominio que requerirían conocimientos en el ámbito agrícola.

Tabla 16. BT de “Foods”

Level 1	Level 2
Foods	Bakery products
Foods	Beverages
Foods	Confectionery
Foods	Cooking fats
Foods	Cooking oils
Foods	Dietetic foods
Foods	Fermented foods
Foods	Frozen foods
Foods	Genetically modified foods
Foods	Health foods
Foods	Irradiated foods
Foods	Novel foods
Foods	Organic foods
Foods	Prepared foods
Foods	Processed foods
Foods	Refrigerated foods
Foods	Salads
Foods	Seafoods
Foods	Simulated foods
Foods	Snack foods
Foods	Soups
Foods	Value added product
Foods	Fast food
Foods	Street foods
Foods	Take out foods

Elaboración del set de datos de distancias de especialización entre conceptos.

Se solicitó a 21 participantes que evaluaran la distancia de especialización entre conceptos en el conjunto seleccionado de términos de AGROVOC. La muestra incluyó estudiantes universitarios, investigadores y profesores procedentes de diferentes ramas del conocimiento.

Como se explicó, el propósito del experimento no era centrarse en un solo tipo de perfil de usuario ya que esto podría limitar o incluso sesgar los resultados del mismo, por lo que la muestra de los términos AGROVOC seleccionados se formó a partir de aquellos de conocimiento común, con la intención de que todos los participantes pudieran entender perfectamente el concepto y la relación.

Se proporcionó a los participantes toda la jerarquía de los términos seleccionados y un formulario en el que debían indicar su apreciación subjetiva de la distancia de especialización entre pares de términos, usando una escala de 5 puntos, donde 5 representaba la distancia más lejana y 1 la más cercana.

Los resultados aportados por las personas que participaron en el experimento se utilizaron para componer un conjunto de datos que contiene las distancias calculadas entre términos de acuerdo a su apreciación subjetiva.

Para apoyar esta tarea, se proporcionó a los participantes formularios que mostraban toda la jerarquía (como se muestra en la tabla 17) y también otros que contenían sólo los términos al mismo nivel (como se muestra en la tabla 18).

Tabla 17. Formulario mostrando la jerarquía completa (extracto de 11 términos)

Distance	Terms (and hierarchy)				
	Foods	Bakery products			
	Foods	Bakery products	Biscuits		
	Foods	Bakery products	Bread		
	Foods	Bakery products	Cakes		
	Foods	Bakery products	Doughs		
	Foods	Bakery products	Puff paste		
	Foods	Bakery products	Batters		
	Foods	Beverages			
	Foods	Beverages	Alcoholic beverages		
	Foods	Beverages	Alcoholic beverages	Beers	
	Foods	Beverages	Alcoholic beverages	Beers	Ales

Tabla 18. Formulario mostrando solo los términos situados por debajo de Foods → Beverages (extracto de 11 términos)

Distance	Terms (and hierarchy)		
	Foods	Beverages	Alcoholic beverages
	Foods	Beverages	Coffee
	Foods	Beverages	Coffee substitutes
	Foods	Beverages	Fruit juices
	Foods	Beverages	Herbal teas
	Foods	Beverages	Soft drinks
	Foods	Beverages	Tea
	Foods	Beverages	Tea substitutes
	Foods	Beverages	Vegetable juices
	Foods	Beverages	Cocoa beverages
	Foods	Beverages	Mate

Evaluación de la interacción con el usuario

La evaluación se realizó siguiendo un enfoque de prueba A / B en el que se enfrentaban los casos que consideraban la distancia de especialización y los que no lo hacían. Estos se presentaron a 40 personas que no participaron en la parte anterior del experimento. Los participantes tuvieron que elegir entre los casos elaborados según el concepto de distancia de especialización y los casos que no lo consideraban, presentando un total de 5 casos de búsqueda a cada usuario.

Se consideraron dos escenarios en los que la información de distancia de especialización entre conceptos podría ser útil desde la perspectiva del usuario, un escenario basado en una interfaz textual y un escenario basado en una interfaz visual:

1) Escenario basado en texto: Un usuario busca un término según un contexto de búsqueda específico, y se le proponen búsquedas relacionadas (como palabras clave) de acuerdo con la información de distancia de especialización entre conceptos. La figura 15 muestra las búsquedas relacionadas sugeridas que no consideran la distancia (ordenada sólo por la jerarquía y el orden alfabético) y la figura 16 las búsquedas relacionadas sugeridas ordenadas por la información de distancia de especialización entre conceptos.

You are looking for "Beverages", suggested terms would be:
 Alcoholic Beverages, Cocoa beverages, Coffee, Coffee Substitutes, Fruit juices, Herbal Teas, Mate, Soft Drinks, Tea, Tea substitutes.

Figura 15. Escenario basado en una interfaz textual sin emplear la información de distancia de especialización entre conceptos.

You are looking for "Beverages", suggested terms would be:
 Alcoholic Beverages, Fruit juices, Soft Drinks, Vegetable juices, Herbal teas, Cocoa beverages, Coffee, Tea, Mate, Apple juice, Grape juice.

Figura 16. Escenario basado en una interfaz textual empleando la información de distancia de especialización entre conceptos.

2) Escenario de interfaz visual: un usuario navega a través de una jerarquía utilizando una interfaz de visualización colocando los términos de acuerdo con la información de distancia de especialización. La figura 17 muestra la jerarquía que no considera la distancia y la figura 18 la jerarquía considerando la distancia, según los resultados mostrados en la tabla 6: el tamaño de las flechas en la figura 4 que conecta los términos representa el concepto de distancia de especialización, proporcional a los valores en la tabla 19.

Tabla 19. Distancia calculada para los términos situados por debajo de Foods → Beverages (extracto de los 11 términos mostrados en la tabla 18)

Terms (and hierarchy)			Distance
Foods	Beverages	Alcoholic beverages	2,83333333
Foods	Beverages	Coffee	4
Foods	Beverages	Coffee substitutes	4,5
Foods	Beverages	Fruit juices	3
Foods	Beverages	Herbal teas	3,16666667
Foods	Beverages	Soft drinks	3
Foods	Beverages	Tea	4
Foods	Beverages	Tea substitutes	4,5
Foods	Beverages	Vegetable juices	3
Foods	Beverages	Cocoa beverages	3,83333333
Foods	Beverages	Mate	3,83333333

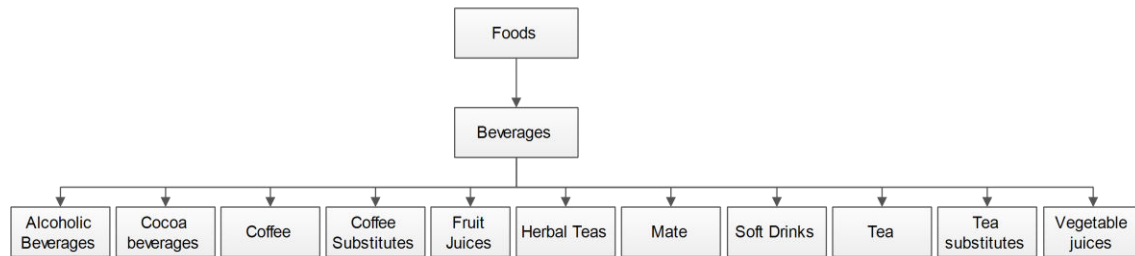


Figura 17. Escenario basado en una interfaz visual sin emplear la información de distancia de especialización entre conceptos.

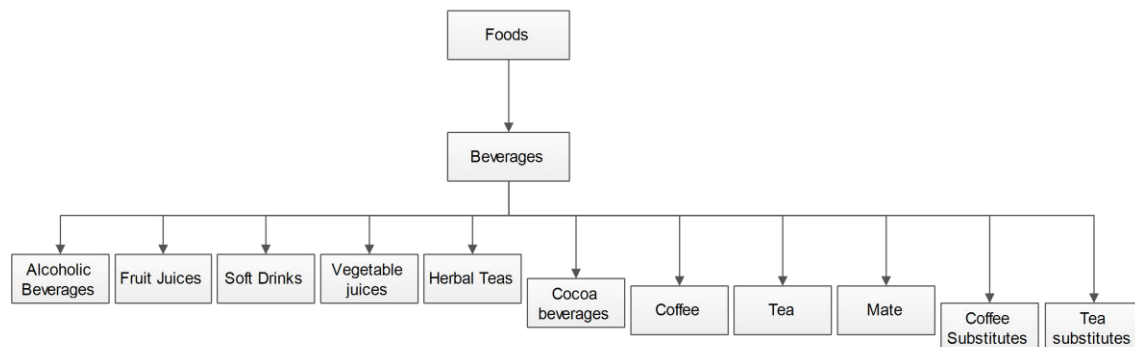


Figura 18. Escenario basado en una interfaz visual empleando la información de distancia de especialización entre conceptos.

No se dio explicación a los participantes sobre la lógica en cuanto al orden de los términos o la longitud de las flechas, puesto que precisamente uno de los aspectos considerados para el experimento fue observar si los participantes podrían reconocer esta lógica por sí mismos.

Con el fin de proporcionar alguna información cualitativa adicional, se entrevistó a todos los participantes después del experimento A / B, pidiéndoles que respondieran a dos preguntas específicas y proporcionaran más explicaciones sobre sus decisiones. Las preguntas específicas fueron:

- ¿Entiende usted la lógica detrás de los casos? (Si no lo hizo se explicó al participante en este momento)
- ¿Considera que el concepto de distancia de especialización es útil (una vez que lo sabe)?

3.4.4 Discusión de los resultados

Resultados de la elaboración del set de datos de distancia de especialización entre conceptos

En la primera parte del experimento se pidió a 21 participantes que evaluaran el concepto de distancia de especialización entre el conjunto de términos AGROVOC anidados bajo "Alimentos" usando una escala de cinco puntos (donde 5 representaban la distancia más lejana y 1 la más cercana).

Los resultados se recogieron en una tabla y se calcularon los valores de media y desviación para cada término, para cada nivel jerárquico y para toda la selección de términos. La tabla 19 contiene un extracto de la tabla de resultados para los términos de nivel 3 y la tabla 20 muestra los resultados globales.

Tabla 20. Extracto de la tabla de resultados para los términos de nivel de profundidad 3.

Terms (and hierarchy)			1	2	3	17	20	21	AVG	DEV
foods	Bakery products	Batters	3	3	3	...	3	3	2,8571	0,3586
foods	Beverages	Alcoholic beverages	3	2	3	3	3	3	2,9048	0,3008
foods	Beverages	Coffee	4	4	4	4	4	4	3,9048	0,3008
foods	Beverages	Coffee substitutes	5	4	4	5	5	4	4,4286	0,5071
foods	Beverages	Fruit juices	3	3	3	3	3	3	3,0952	0,3008
foods	Beverages	Herbal teas	3	3	3	3	4	3	3,1429	0,3586
foods	Beverages	Soft drinks	3	3	3	3	3	3	3,0952	0,3008
foods	Beverages	Tea	4	4	4	4	4	4	3,9048	0,3008
foods	Beverages	Tea substitutes	5	4	4	5	5	4	4,4286	0,5071
foods	Beverages	Vegetable juices	3	3	3	3	3	3	3,0952	0,3008
foods	Beverages	Cocoa beverages	4	4	3	4	4	4	3,6190	0,4976
foods	Beverages	Mate	3	4	4	4	4	4	4,0476	0,3842
foods	Confectionery	Cakes	3	3	3	3	3	3	2,9524	0,2182
foods	Confectionery	Chewing gum	4	4	4	4	4	4	3,9048	0,3008
foods	Confectionery	Sugar confectionery	3	3	3	3	3	3	2,9048	0,3008
foods	Confectionery	Chocolate	3	4	3	3	4	4	3,3333	0,4830

Tabla 21. Resultados globales del set de datos de la distancia de especialización entre conceptos.

Level	#Terms	Distance (AVG)	Deviation (AVG)
1 --> 2	25	3,820952381	0,443057077
2 --> 3	47	2,80952381	0,402373908
3 --> 4	22	1,413419913	0,393588839
4 --> 5	15	1,038095238	0,095618289
5 --> 6	4	1	0
TOTALS	113	2,654135338	0,353922064

Como se puede ver en la tabla 21, tanto el acuerdo de calificación global como el realizado por nivel de profundidad son significativamente altos, además, la desviación para cada término fue inferior a 1.

Adicionalmente, a cada usuario se le preguntó acerca de la adecuación de usar una escala de cinco puntos para la calificación, si tenían problemas debido al rango de escala o si se sentirían más cómodos usando otro sistema. Los 21 participantes declararon que consideraban adecuado usar una escala de cinco puntos para la selección de términos. Once de ellos (52,3%) subrayaron que en el caso de utilizar una escala mayor (por ejemplo, decimal o una escala de siete puntos) tendrían problemas para asignar valores intermedios. Ocho de ellos (38%) mencionaron que para otros casos con un mayor número de términos podría ser necesario utilizar una escala más grande, pero para ello se necesitaría un conocimiento de dominio adecuado para poder asignar valores precisos usando una mayor escala.

Durante el experimento, varios participantes señalaron algunas incoherencias en el tesauro de AGROVOC; por ejemplo, consideraron que el "Mate" no debería estar al mismo nivel jerárquico que "Herbal Teas" y un caso similar que involucra "Tea", "Coffe", "Tea substitutes" y "Coffee substitutes".

También se sorprendieron e incluso se molestaron por el número de términos en ciertas categorías, por ejemplo, una pregunta recurrente fue que cómo podría haber sólo 4 términos anidados en "Food → Seafood" y cinco en "Foods → Prepared foods → Soyfoods".

Este tipo de comentarios mostraron cómo el proceso de evaluación no sólo fue útil para preparar el conjunto de datos, (dado que, aunque esta parte del experimento fue inicialmente planeada y diseñada sólo lo fue para elaborar un conjunto de datos con valores relativos a la distancia de especialización entre conceptos), sino que, gracias a la información cualitativa obtenida de los usuarios potenciales, se evidenció que este tipo de evaluación podría ser una herramienta útil para localizar y analizar inconsistencias en un KOS.

Resultados de la evaluación de la interacción con el usuario

Como se describió anteriormente, se concibieron dos escenarios en los que la medida de la distancia de especialización entre conceptos podría ser útil para la recuperación de información desde la perspectiva de usuario final. La primera fue orientada a una interfaz textual (figuras 15 y 16) y la segunda para el uso de una interfaz visual básica tipo árbol (figuras 17 y 18). La tabla 22 muestra el número y el porcentaje de participantes que prefirieron las interfaces utilizando el concepto de información de distancia de especialización.

Tabla 22. Número y porcentaje de participantes que prefirieron las interfaces que empleaban la distancia de especialización entre conceptos.

Case	Text-based scenario using concept specialization distance		Visual-interface scenario using concept specialization distance	
	numbers	percentage	numbers	percentage
A	35	88%	22	55%
B	34	85%	26	65%
C	35	88%	33	83%
D	33	83%	24	60%
E	25	63%	24	60%

Como se puede ver en la tabla 22, aunque hubo más participantes que prefirieron la interfaz visual que empleaba la información de distancia de especialización entre conceptos, los números no eran realmente significativos, excepto en el caso C que excedía el 80%. Por otro lado, para los mismos casos en el escenario basado en el texto, el número de participantes que prefirió los términos sugeridos seleccionados haciendo uso de la información de distancia de especialización entre conceptos, fue significativamente mayor.

Después de la evaluación se preguntó a todos los participantes acerca de su comprensión de la lógica de los casos. Una vez que todos los participantes entendieron la lógica, se les preguntó si encontraban útil la información en cuanto a la distancia de especialización entre conceptos para fines de recuperación de información como los mostrados en el experimento.

Diecinueve participantes (47,5%) fueron capaces de explicar la lógica detrás de los casos utilizando el concepto de distancia de especialización; sin embargo, la mayoría de ellos sólo se dio cuenta después de ver los dos o tres primeros casos; mientras que todos ellos se dieron cuenta del orden alfabético en el primer caso.

Cuando se explicó la lógica, 25 participantes (62,5%) consideraron útil el uso de la distancia de especialización en la gran mayoría de los casos. Once de ellos (27,5%) argumentaron que podría ser útil, pero no para todos los casos; particularmente encontraron que podría retrasar a los usuarios cuando tienen una idea clara sobre lo que están buscando, puesto que les resultaba más fácil encontrar un término cuando los términos se ordenan alfabéticamente. Por último, 4 participantes (10%) explicaron que preferían mayoritariamente el orden alfabético y, además, dos de estos 4 participantes (5%) mantuvieron esta preferencia incluso cuando pudieron darse cuenta de la lógica detrás del concepto de distancia de especialización sin explicación previa.

La razón de la diferencia relacionada con el número de participantes que prefiere el enfoque basado en texto utilizando el concepto de distancia de especialización, pero no la interfaz visual para el mismo caso, parece estar en la propia interfaz. Existen múltiples maneras de visualizar un KOS y emplear una visualización u otra para un KOS en particular puede afectar dramáticamente al usuario final (Gaona et al., 2014).

Como se puede ver en las figuras 17 y 18, para el experimento se eligió una estructura tipo árbol muy simple para representar la jerarquía de KOS y luego se empleó la longitud de las flechas para representar la distancia de especialización. De esta manera, los términos en el mismo nivel de jerarquía no fueron colocados a la misma altura, lo que resultó inquietante y caótico para varios participantes, especialmente cuando eran incapaces de darse cuenta de la lógica detrás de este enfoque, en contraste con los elementos del otro caso, los cuales estaban colocados siguiendo un orden claro. Por otra parte, como ya se ha dicho, una vez entendida o explicada la lógica, la mayoría de los participantes consideraron que el enfoque de la distancia de especialización entre conceptos podría resultar útil tanto para las interfaces visuales como basadas en texto.

3.4.5 Conclusiones del experimento

El objetivo principal de este experimento fue evaluar las oportunidades de la distancia de especialización entre conceptos desde la perspectiva del usuario final. Por esta razón, el primer paso fue construir un conjunto de datos de estas distancias, de acuerdo con los valores proporcionados por usuarios potenciales en lugar de utilizar expertos de dominio para asignarlos.

Los resultados mostraron un alto nivel de acuerdo, lo que permitió construir un set de datos adecuado para la segunda parte del experimento. Además, la metodología resultó ser útil para detectar inconsistencias en el KOS y puede ayudar en el control de calidad y optimización de las relaciones jerárquicas. No se ha analizado más este aspecto ya que no formaba parte del

experimento ni se contemplaba durante su planificación y diseño, pero la información obtenida mediante el uso de esta metodología muestra claramente sus oportunidades.

La evaluación de la interacción del usuario final siguió un enfoque de prueba A / B proponiendo casos de uso implementados a través de interfaces tanto visuales como basadas en texto. Los primeros resultados mostraron un gran potencial para la interfaz basada en texto, mientras que las expectativas para la interfaz visual permanecieron más bajas.

Las entrevistas con los participantes reflejaron que el enfoque de visualización empleado para el experimento no era la opción más adecuada para mostrar plenamente el potencial de la integración de la información de distancia de especialización entre conceptos en este tipo de interfaces. La mayoría de los participantes encontraron que el enfoque podría resultar útil tanto para interfaces visuales como basadas en texto, una vez que se les explicó la lógica de la construcción de la interfaz.

Como las posibilidades de uso para interfaces basadas en texto parece estar claro, el trabajo futuro debería centrarse en utilizar la información de distancia de especialización entre conceptos para construir diferentes interfaces con el fin de encontrar la mejor manera de integrarlo. El conjunto de datos elaborado durante este experimento podría ser utilizado para este propósito. Otra de las líneas de investigación futuras abarcaría el uso de la distancia de especialización entre conceptos, no sólo para el término anterior o posterior de la jerarquía, sino también entre términos situados a la misma profundidad.

Capítulo 4: Resumen de los experimentos realizados en cuanto a interfaces visuales con KOS

“You must be shapeless, formless, like water. When you pour water in a cup, it becomes the cup. When you pour water in a bottle, it becomes the bottle. When you pour water in a teapot, it becomes the teapot. Water can drip and it can crash. Become like water my friend.”

(Bruce Lee, en 1971)

4.1 Motivación

Al indagar sobre el limitado uso de los KOS respecto a sus posibilidades teóricas, la causa más elemental que podría preverse estaría ligada a problemas de usabilidad en la interfaz y a una visualización inadecuada. De hecho, la motivación de la presente tesis doctoral se origina precisamente en el estudio de las interfaces de Organic.Edunet, donde se observaron factores que dificultaban el uso de los KOS que no estaban ligados a la interfaz sino a las características del propio KOS.

En lo que afecta a las preguntas de investigación formuladas, la realización de experimentos ligados a interfaces visuales con KOS, de forma paralela a los descritos en el capítulo 3, que constituyen el núcleo central de la tesis; ha resultado de especial utilidad para proporcionar evidencias adicionales sobre los problemas de usabilidad de los KOS y las dificultades que esto origina para su integración en interfaces.

El capítulo actual resume el trabajo realizado en los estudios más relevantes, centrándose en los aspectos que respaldan la investigación desarrollada en el capítulo 3 y aquellos que facilitan la comprensión de las conclusiones finales de la presente tesis.

4.2 Introducción

En *“Applying visualization techniques to develop interfaces for educational repositories: the case of Organic.Lingua and VOA3R.”* (Martín-Moncunill, Sánchez-Alonso, Gaona-García, & Marianos, 2013), se analiza la integración de interfaces visuales que empleaban KOS en los repositorios VOA3R y Organic.Edunet.

En este trabajo se observaron diversos problemas de usabilidad en las interfaces, así como otros problemas que habría que abarcar antes de tratar específicamente con la usabilidad de la interfaz; a saber, la percepción de la utilidad por parte del usuario, la propia usabilidad del KOS, el entendimiento de su estructura y la necesidad de un tiempo de aprendizaje a consecuencia de esto.

Este artículo se realizó mientras se llevaba a cabo otro estudio relacionado de mayor alcance “*A usability study of taxonomy-visualization user interfaces in digital repositories*” (Gaona García, Martín-Moncunill, Sánchez-Alonso, & García, 2014), publicado poco después, el cual realizaba un estudio de usabilidad comparando las interfaces para la representación de KOS más ampliamente utilizadas, sobre el tesoro AAT (The Getty Research Institute, 2017), empleado en Europeana (Purday, 2009).

En el artículo “*Selection and Use of Search Mechanisms in Learning Object Repositories: the Case of Organic.Edunet*” (Martin-Moncunill, Gaona-García, Garcia-Barriocanal, & Sanchez-Alonso, 2015) se resumen los resultados de un estudio de usabilidad llevado a cabo en el portal Organic.Edunet donde se solicitó a los usuarios realizar diferentes tareas de búsqueda que estaban claramente orientadas al uso de una de las interfaces existentes y se proporciona una instantánea sobre cómo se transfirió este conocimiento al proyecto Organic.Lingua; la evolución del portal Organic.Edunet.

En 2016 se publican dos artículos que profundizan sobre el uso de KOS centrados en interfaces de realidad virtual, estos son “*User Experiences in Virtual Reality Environments Navigation Based on Simple Knowledge Organization Systems*” (Martin-Moncunill, Gaona-García, Gordillo-Orjuela, & Montenegro-Marin, 2016) y “*Navigation and Visualization of Knowledge Organization Systems using Virtual Reality Glasses.*” (Gaona-García, Martin-Moncunill, Gordillo-Orjuela, & González-Crespo, 2016). En estos dos experimentos se trabajó con un KOS elaborado ad-hoc, que representaba la taxonomía de animales de un parque zoológico.

Finalmente, a comienzos del 2017 se publica “*Trends and challenges of visual search interfaces in digital libraries and repositories*” (Gaona-García, Martín-Moncunill, & Montenegro-Marin, 2017), un análisis sistemático de la literatura, donde se abordan las principales tendencias y retos en cuanto a interfaces de visualización implementadas sobre KOS, profundizando en el contexto de las librerías y repositorios digitales. Las siguientes secciones resumen brevemente cada uno de los experimentos.

4.3 Applying visualization techniques to develop interfaces for educational repositories: the case of Organic.Lingua and VOA3R

Este estudio parte de la base de que la mayoría de los repositorios educativos incluyen algún tipo de KOS con el cual los responsables de estos repositorios catalogan sus recursos (Stefaner, y otros, 2009). Por otro lado, considera que las representaciones visuales como los diagramas han demostrado sobradamente su efectividad en el área educativa por lo que emplear este tipo de técnicas visuales para navegar por KOS, resultaría de utilidad al menos a nivel teórico.

Sin embargo, los datos de uso de las interfaces visuales implementadas en el ámbito de los proyectos Organic.Edunet (Sánchez-Alonso & Sicilia, Using an AGROVOC-based ontology for the description of learning resources on organic agriculture , 2009) y VOA3R (Solodovnik, 2013) no reflejaba el interés esperado por estas interfaces. Con tal motivo se realizó un análisis de usabilidad de las interfaces navegacionales de estos repositorios; la figura 19 muestra la interfaz de Organic.Edunet y la figura 20 muestra la interfaz de VOA3R.

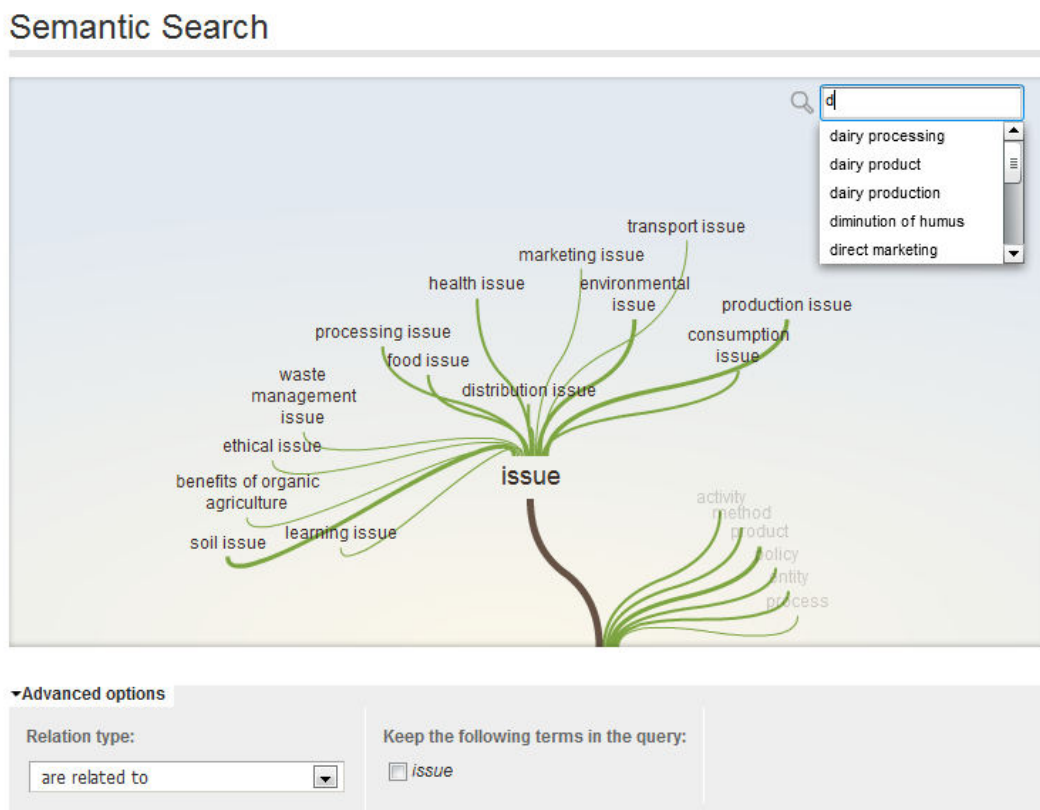


Figura 19. Interfaz navegacional de Organic.Edunet

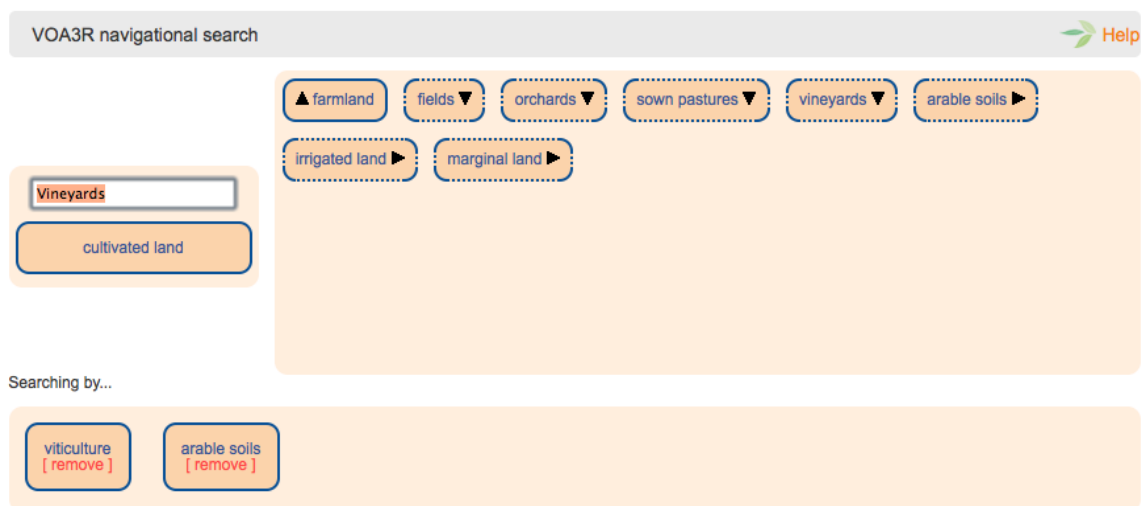


Figura 20. Interfaz navegacional de VOA3R

Los estudios, que quedaron plasmados en los entregables de ambos proyectos de investigación (Stracke, Sicilia, Helmstedt, Martín-Moncunill, & Protonotarios, 2013), resultaron de utilidad para identificar diversos problemas de usabilidad, como los ilustrados en la figura 21.

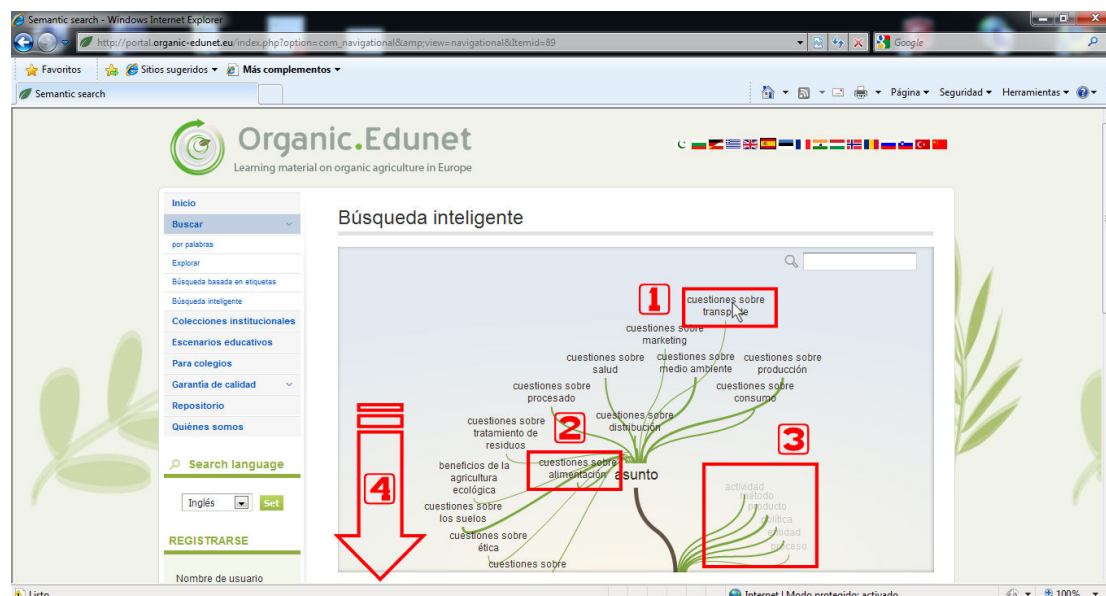


Figura 21. Interfaz navegacional de VOA3R

En esta figura aparecen marcados problemas relativos a la incorrecta identificación de enlaces (1), término montado sobre la interfaz por ausencia de espacio (2), problemas de contraste (3) o no percatarse de la posibilidad de hacer *scroll* y ver los términos que se cargan en la zona que queda fuera de pantalla.

Por encima de los problemas de usabilidad de la interfaz que se detectaron a nivel individual, los estudios realizados apuntaron hacia tres problemas que afectaban a las dos interfaces:

- Utilidad: Los participantes no eran capaces de apreciar la utilidad de las interfaces; esto es, en qué forma estas les resultarían útiles para llevar a cabo las tareas.
- Usabilidad del KOS, problemas en el entendimiento del esquema de clasificación.
- Necesidad de dedicar tiempo a aprendizaje. Con motivo de lo anterior, para emplear la interfaz, los usuarios requerirían un tiempo para entender el KOS. Nótese que esto va más allá del factor de usabilidad ligado con la facilidad de aprendizaje de la interfaz.

4.4 A usability study of taxonomy-visualization user interfaces in digital repositories

El análisis llevado a cabo en este estudio analiza diferentes enfoques para mostrar la información en un repositorio de recursos digitales. Para ello se realizó una selección de técnicas de acuerdo a estructuras de representaciones gráficas (Herman, Melancon, & Marshall, 2000), estrategias de *clustering* temático (Merčun, Žumer, & Aalberg, 2012) y, finalmente, a través de representaciones jerárquicas (Shneiderman & Plaisant, 2006). Basándose en esta selección se desarrollaron ocho tipos de interfaces inspiradas en 8 técnicas de visualización, clasificadas en cuatro estrategias: radial, hiperbólica, tipo árbol y por categorías. Estas interfaces se representan en la figura 22.

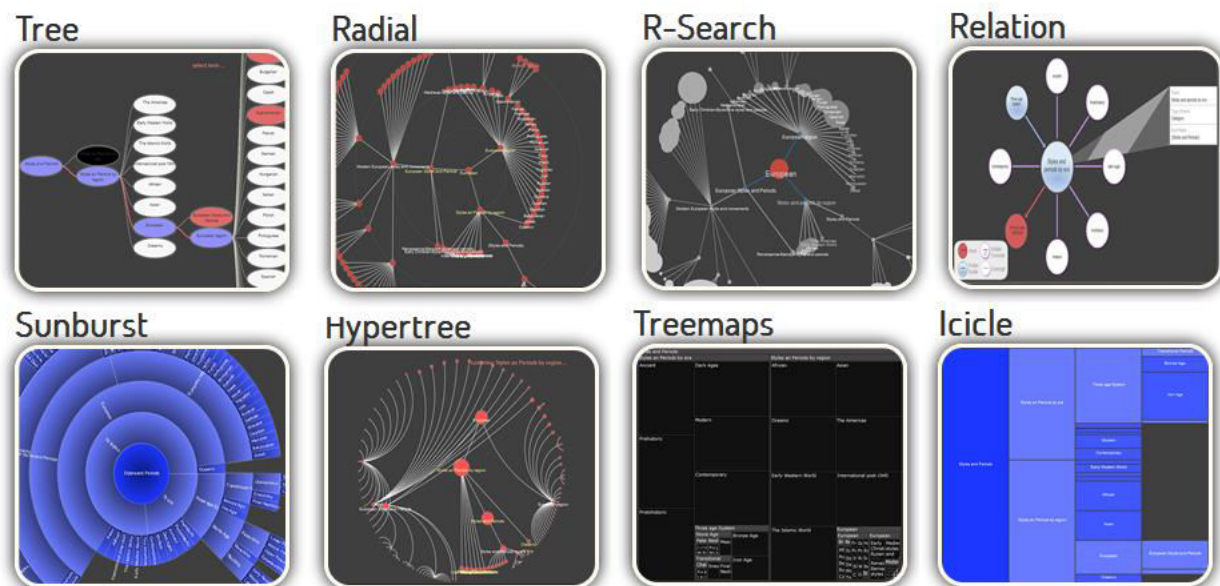


Figura 22. Las ocho interfaces visuales desarrolladas para el estudio

Las herramientas visuales instaladas en repositorios digitales deberían idealmente ayudar a los usuarios a ubicar los recursos digitales de forma eficaz y precisa de acuerdo con una estructura temática o área de conocimiento, así como dar una idea de la cobertura temática del repositorio. Esta idea de la cobertura temática se refiere a que la interfaz gráfica debe indicar de manera sencilla cómo está cubierto, en cuanto al número de recursos se refiere, un tema concreto.

El análisis se dividió en 2 partes, en la primera se analizaron las habilidades perceptivas para usar interfaces (atención, retención de información y comprensión), configurando las ocho interfaces para trabajar con KOS muy simples, empleando términos de conocimiento común para evitar el sesgo producido por la complejidad de los mismos. La figura 23 muestra una de las interfaces empleadas para esta parte del experimento, en una disposición radial y tratando la temática de tipos de deportes. Esta figura 23 también representa un ejemplo de cómo mostrar la cobertura temática empleando círculos o nodos de diferentes tamaños.



Figura 23. Una de las interfaces empleadas para la primera parte del estudio. El tipo de visualización de la interfaz es radial, el tamaño de los círculos representa la cobertura.

La segunda parte, más centrada en la evaluación de la usabilidad utilizó una colección simulada construida por 40.000 recursos digitales de Europeana, clasificándolos de acuerdo a la rama de conocimiento de “Estilos y Períodos” del Tesauro de Arte y Arquitectura (AAT). La figura 24 muestra la representación de parte de esta rama del conocimiento empleando la interfaz denominada como “*Sunburst*”, los sectores muestran la cobertura temática, los números sobreimpresos (no existen en la interfaz original) indican el nivel de profundidad.

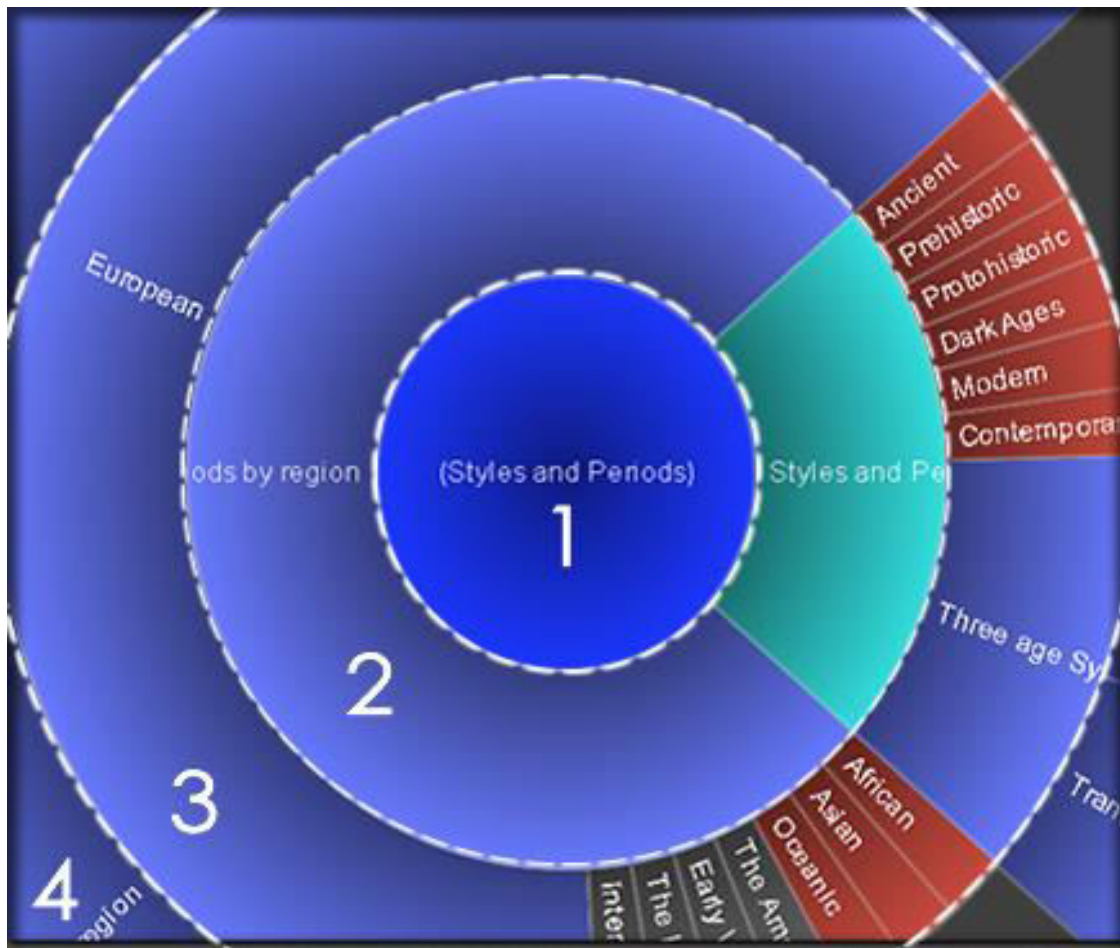


Figura 24. Una de las interfaces empleadas para la segunda parte del estudio, denominada como “Sunburst”, también de tipo radial, el tamaño de los sectores representa la cobertura.

Los resultados obtenidos en el estudio apuntaron una vez más a la utilidad teórica de estas interfaces, pero cabe destacar las variaciones en resultados tanto cualitativos como cuantitativos que existieron para algunas interfaces al cambiar el KOS, hasta el punto que algunos usuarios llegaron a considerar que no se trataba de la misma interfaz. Esto reforzó la idea de focalizarse en la utilidad y usabilidad del KOS en sí mismo y en la necesidad de adaptar convenientemente KOS e interfaz de acuerdo a los perfiles de usuario, teniendo en mente el contexto de uso y objetivos específicos de los usuarios para los que se desarrollara la interfaz.

4.5 Selection and Use of Search Mechanisms in Learning Object Repositories: the Case of Organic.Edunet

Entre los objetivos del proyecto europeo de investigación del repositorio de objetos educativos centrado en la agricultura ecológica Organic.Edunet, se encontraba la investigación sobre interfaces de búsqueda innovadoras (Stoitsis, Protonotarios, Sanchez-Alonso, Martin Moncunill, & Kastrantas, 2013), incluyendo aquellas que sacaran mayor provecho a los sistemas de organización del conocimiento y la información semántica (Palavitsinis, Manouselis, & Sánchez-Alonso, Evaluation of a metadata application profile for learning resources on organic agriculture, 2009) que se emplearon en el proyecto.

Siguiendo este objetivo y las recomendaciones de los revisores de la Comisión Europea, se implementaron distintas interfaces de búsqueda con la intención de facilitar la localización de sus recursos. Considerando que atendían a los objetivos y contexto específico de la búsqueda, sería más conveniente para los usuarios emplear una de ellas sobre el resto.

Con tal motivo, al finalizar el proyecto y en vistas a la mejora del portal, dentro del marco del proyecto Organic.Lingua (destinado fundamentalmente a la incorporación de servicios multilingües) se desarrolló un estudio de usabilidad que comparaba los resultados de las distintas interfaces e indagaba sobre la elección de los usuarios entre las interfaces existentes a la hora de realizar una búsqueda.

Para ello se realizaron test de usabilidad (Rubin & Dana, 2008) donde a los usuarios se les proponían distintas tareas de búsqueda, quedando a elección de los mismos qué interfaz emplear para llevar a cabo la tarea. Las tareas no forzaban estrictamente la elección de una interfaz de búsqueda determinada, pero para cada tarea de búsqueda existían una o varias interfaces que se consideraba resultaban más adecuadas.

Así, los usuarios podrían completar la tarea empleando otras interfaces, pero igualmente se esperaba que estos tuvieran que invertir más tiempo o/y cometer más errores. La tabla 23 muestra un ejemplo de esto para tres tareas:

Tabla 23. Tareas de búsqueda “C”, “E” y “H” y métodos más adecuados para llevarlas a cabo.

TAREA	MÉTODOS ADECUADOS
C) Accede a un recurso relacionado con el proceso de elaboración de productos lácteos en cualquier idioma.	Localizar el concepto “elaboración de productos lácteos” en la “ <i>Búsqueda Semántica</i> ”. También puede ser completada con la “ <i>Búsqueda por Palabras</i> ” pero el usuario debe comprobar que esa es la temática, no que simplemente aparezcan las palabras.
E) ¿Cuántos recursos podrías asegurar que tratan sobre la elaboración de productos lácteos?	La “ <i>Búsqueda Semántica</i> ” es la única que puede garantizar esta relación.
H) El tema del impacto ambiental es muy tratado en la web de Organic.Edunet. Averigua cuantos recursos hay al respecto en todos los idiomas.	El término se puede localizar directamente en la “Nube de Conceptos”. También se puede completar empleando la “ <i>Búsqueda Semántica</i> ” dado que “impacto ambiental” es un término de la ontología de Organic.Edunet.

Tal como muestra la tabla 24, para cada usuario (#U) se tomaron datos sobre el número de errores (ERR), el tiempo empleado (“TIME”), las interfaces empleadas (columnas desde “PA” hasta “OT”), si el usuario requirió ayuda (“A”), si tuvo éxito en completar la tarea (“E”) y con qué interfaz lo logró (representada con una X en negrita y subrayada). La tabla 24 muestra estos resultados para la tarea “C”.

Tabla 24. Resultados para la tarea “C”

TAREA C										
#U	ERR	TIME	PA	EX	SE	ET	NU	OT	A	E
1	7	5' 07"	X	<u>X</u>	-	-	-	X	N	S
2	4	1' 54"	-	-	-	-	<u>X</u>	-	N	S
3	2	1' 41"	X	-	-	X	-	-	N	N
4	5	3' 41"	-	-	<u>X</u>	-	-	-	S	S
5	10	5' 37"	X	X	X	-	-	-	N	N
6	10	6' 18"	X	-	X	-	X	-	N	N
7	1	2' 05"	<u>X</u>	-	-	-	-	-	N	S
8	5	3' 59"	X	X	<u>X</u>	X	-	-	N	S
9	6	5' 58"	X	-	<u>X</u>	X	-	-	S	S
10	5	3' 21"	X	-	-	-	-	-	N	S
11	4	2' 30"	-	-	<u>X</u>	-	-	-	N	S
12	5	3' 24"	<u>X</u>	-	-	-	X	-	N	S
13	5	3' 21"	-	-	<u>X</u>	-	-	-	S	S

El trabajo se centró en analizar cómo orientaron sus búsquedas los sujetos del estudio, si se adecuaban a lo que teóricamente correspondía a la opción óptima y cuáles fueron las interfaces que los usuarios intentaron emplear sin éxito. La figura 25 muestra en cuantas ocasiones se intentó emplear una interfaz para completar una tarea y su tasa de éxito, mientras que las figuras 26 y 27 muestran gráficamente los resultados de la tabla 24.

Para cada tarea se analizaron en profundidad tanto los datos cuantitativos como los cualitativos, que incluyeron opiniones vertidas por los usuarios durante la sesión de test, cuestionario post-test y entrevistas, teniendo además en cuenta los datos recogidos en el formulario empleado para la selección de los participantes.

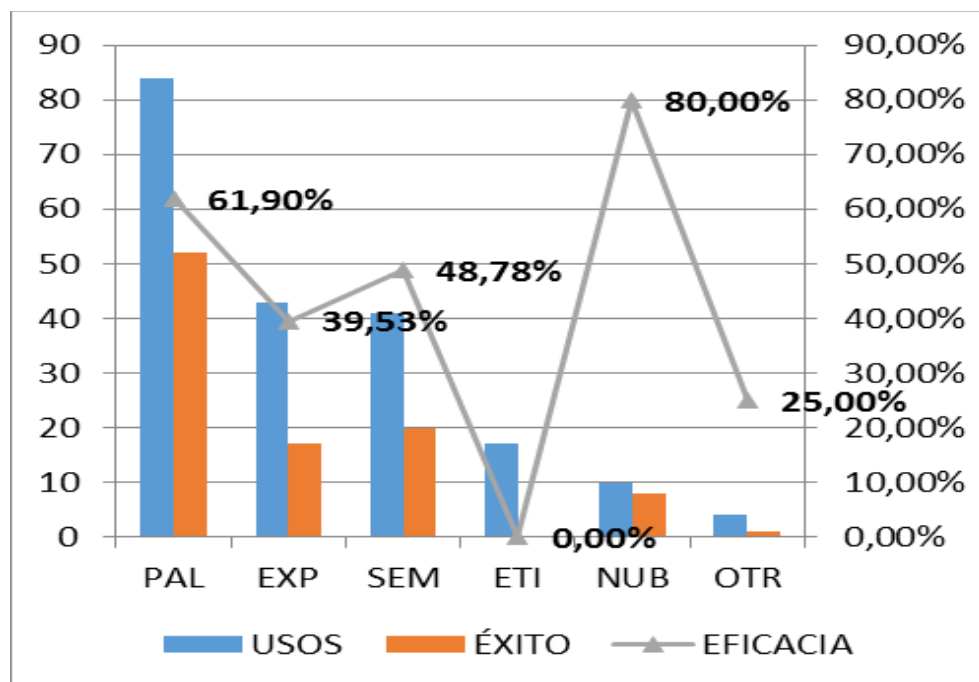


Figura 25. Ocasiones en las que se intentó emplear una interfaz para completar una tarea y su tasa de éxito.

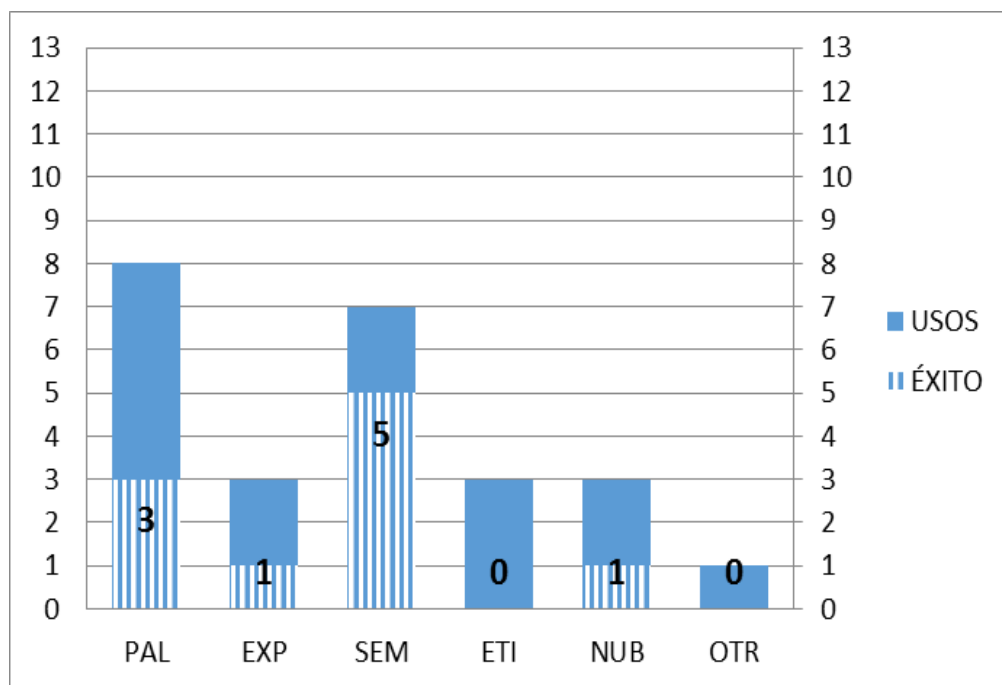


Figura 26. Tasa de éxito para cada interfaz de búsqueda en la tarea C.

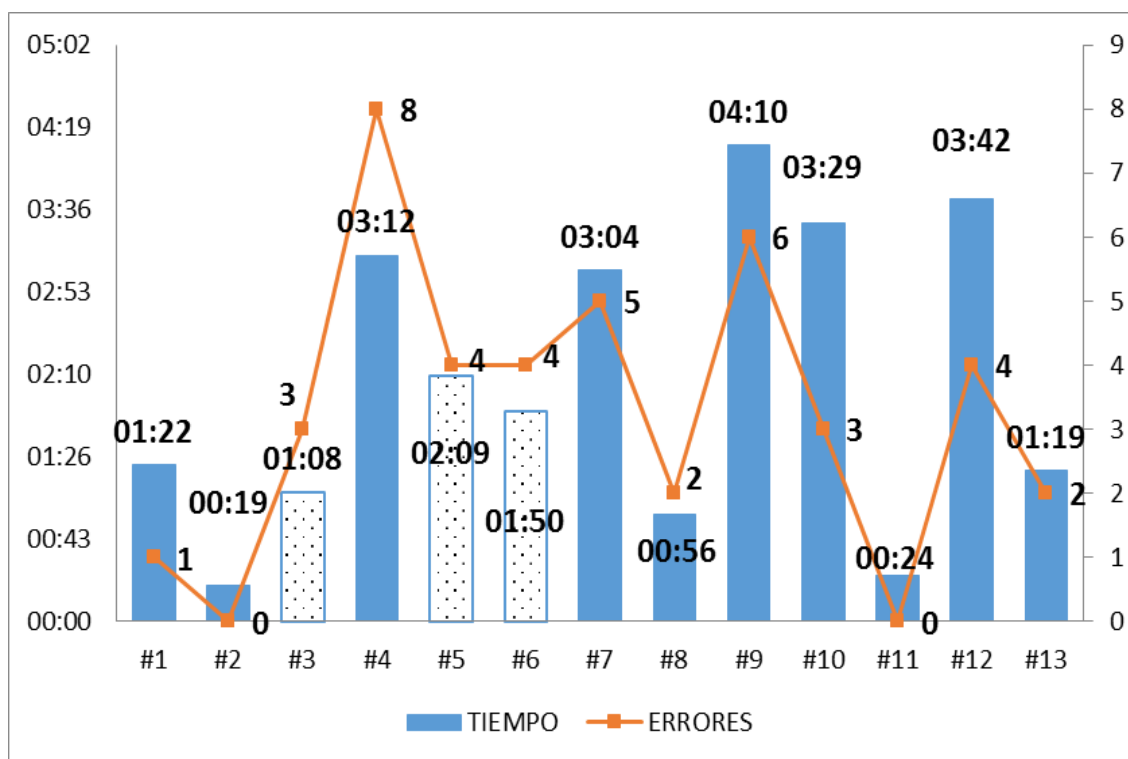


Figura 27. Tiempos y errores en la tarea C para cada usuario (las columnas con relleno punteado indican que la tarea no se completó satisfactoriamente).

El estudio llevado a cabo en el portal de Organic.Edunet, además de permitir detectar un gran número de problemas de usabilidad específicos para cada interfaz, permitió indagar sobre la causa del desuso de estas interfaces y el proceso de decisión de los usuarios para elegir un tipo de interfaz sobre otra en sus búsquedas.

Los resultados del mismo mostraron claramente que el número de interfaces de búsqueda existentes en el portal, a las que se accedía a través de un menú principal, acabó constituyendo un problema de usabilidad en sí mismo, puesto que los usuarios no eran capaces de percibir qué interfaz les sería más útil para realizar búsquedas.

La idea de poner a disposición del usuario múltiples interfaces de búsqueda resulta completamente inadecuada, especialmente cuando los usuarios desconocen la utilidad de cada interfaz. Por ello, la orientación a la hora de implementar la nueva interfaz dentro del proyecto Organic.Lingua se centró en combinar el potencial de las distintas interfaces en una interfaz de búsqueda única de la forma más transparente posible de cara al usuario, ofreciéndole la información y funcionalidad necesaria en el momento oportuno. La figura 28 muestra las recomendaciones de desarrollo para la interfaz de Organic.Lingua.

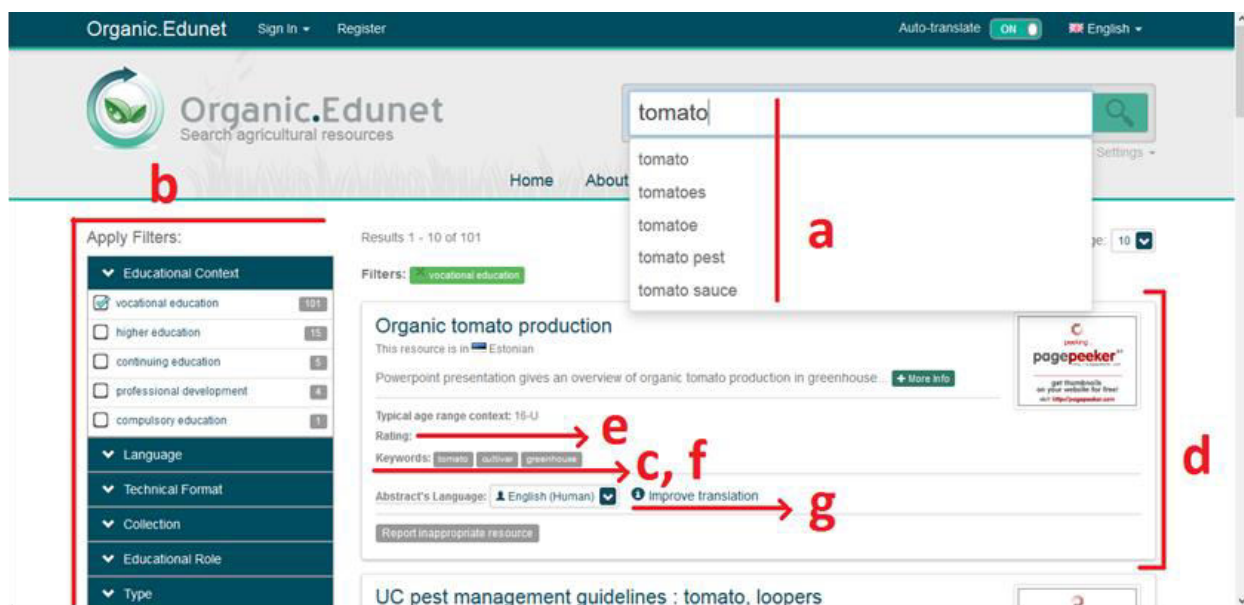


Figura 28. Recomendaciones de desarrollo para la interfaz de Organic.Lingua.

De esta interfaz destacan las siguientes funcionalidades:

- El cuadro de búsqueda reconoce automáticamente etiquetas y términos, como los conceptos que aparecían en la interfaz de “nube de conceptos” presente en Organic.Edunet y los sugiere. (A)
- Mediante el uso de filtros y búsqueda por facetas, se consigue que interfaces como la de “Exploración”, presente en Organic.Edunet y específica para realizar búsquedas por categorías tipo “directorio” no se haga necesaria para este caso. (B)
- En los resultados es posible hacer “click” sobre el término/etiqueta y mostrará todos los que estén relacionados con el mismo, haciendo uso de las relaciones recogidas en la interfaz semántica de Organic.Edunet (C).
- Para cada resultado se muestra la información más relevante atendiendo a lo expresado por usuarios, en diversos estudios que se realizaron durante el proyecto (D).
- Facilidad para que los usuarios califiquen los resultados de búsqueda. (E).
- Los términos que formen parte de un esquema de conocimiento permitirán acceder a una interfaz visual para navegar por el mismo, similar a la búsqueda inteligente de Organic.Edunet, pero dejando claro al usuario la utilidad y contexto de uso adecuado (F).
- Permite que los usuarios puedan sugerir modificaciones y traducciones del título o el resumen de un resultado (G).

4.6 Estudios publicados relacionados con interfaces que emplean realidad virtual

Los últimos estudios realizados se han desarrollado en el ámbito de la realidad virtual. Concretamente se trata de dos estudios preliminares, donde la implementación de los KOS empleados se realizó teniendo en cuenta la necesidad de integrarse en una interfaz visual con características concretas. Recalcar que las particularidades de las interfaces inmersivas que se crean para dispositivos de realidad virtual presentan multitud de retos que no están directamente vinculados con la integración de los KOS y que quedan fuera del alcance de esta tesis.

En el artículo *“Navigation and Visualization of Knowledge Organization Systems using Virtual Reality Glasses: first insights”* (Gaona-García, Martín-Moncunill, Gordillo-Orjuela, & González-Crespo, 2016) se describe el proceso seguido para la creación de un KOS que representaba la taxonomía de animales de un parque zoológico, concretamente el Bio-Parc de Fuengirola, Málaga (España) y la implementación de una interfaz inmersiva de realidad virtual que hacía uso de la misma.

Esta implementación fue desarrollada usando el motor unity 3D, con la idea de emplear objetos usados en escenas de videojuegos para representar el KOS del Bio-Parc con nodos que enmarcaran relaciones jerárquicas como las que trata el capítulo anterior. Para esto se empleó el API de Google Cardboard para el manejo de eventos táctiles y el seguimiento de los movimientos de la cabeza.

La figura 29 muestra la representación simple del KOS en entorno 3D, centrado en los animales presentes en los humedales y la figura 30 presenta una visión más amplia del mismo. La estructura presente en la figura 29 se obtuvo del catálogo de especies proporcionado por el parque.

“Humedales” es un nodo hijo de la zona de África, en este sentido se puede deducir que todas las especies de los “Humedales” pertenecen a África y pueden encontrarse en dicho contexto. La figura 30 muestra toda la estructura del KOS de manera expandida además de la cámara que permite el proceso de navegación dentro de la aplicación; cada zona tiene su propio color y sus nodos hijos tienen colores en una escala similar con la intención de reforzar visualmente su relación con el nodo padre. Por otro lado, las figuras 31 y 32 muestran la representación en el entorno de realidad virtual. La figura 31 presenta una vista más amplia, mientras que la 32 está centrada en un animal concreto.

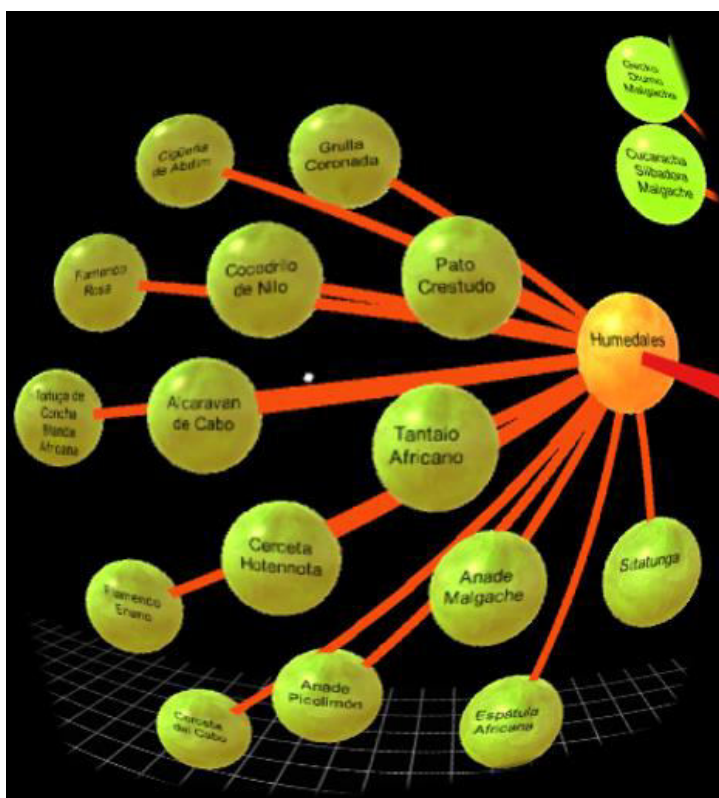


Figura 29. Representación simple del KOS en entorno 3D, centrado en los animales presentes en los humedales.

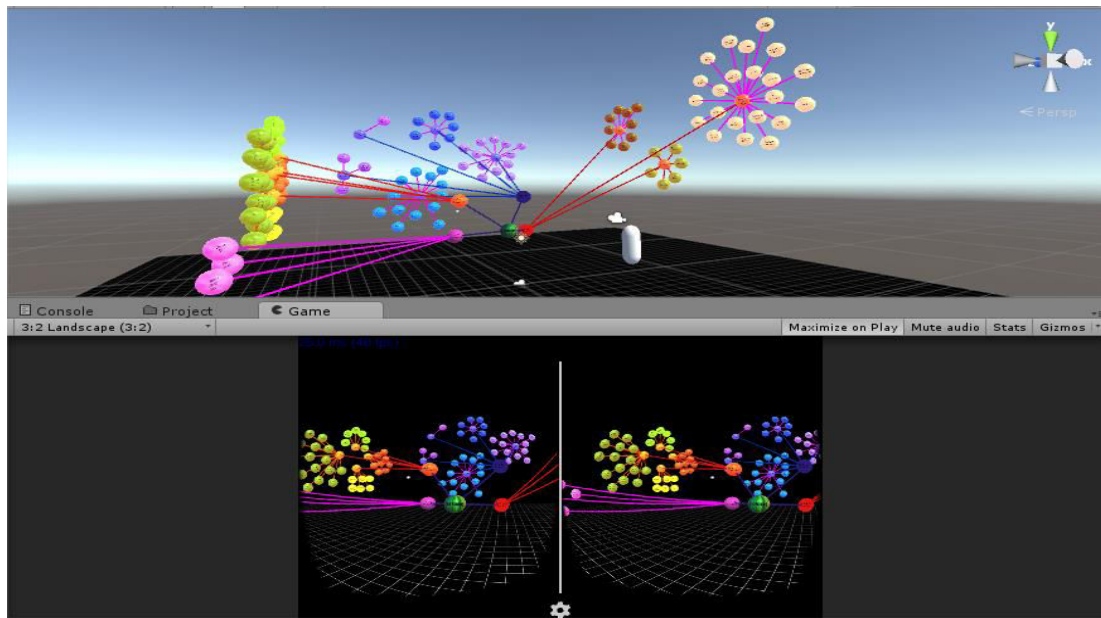


Figura 30. Una visión más amplia del KOS en Unity SDK.

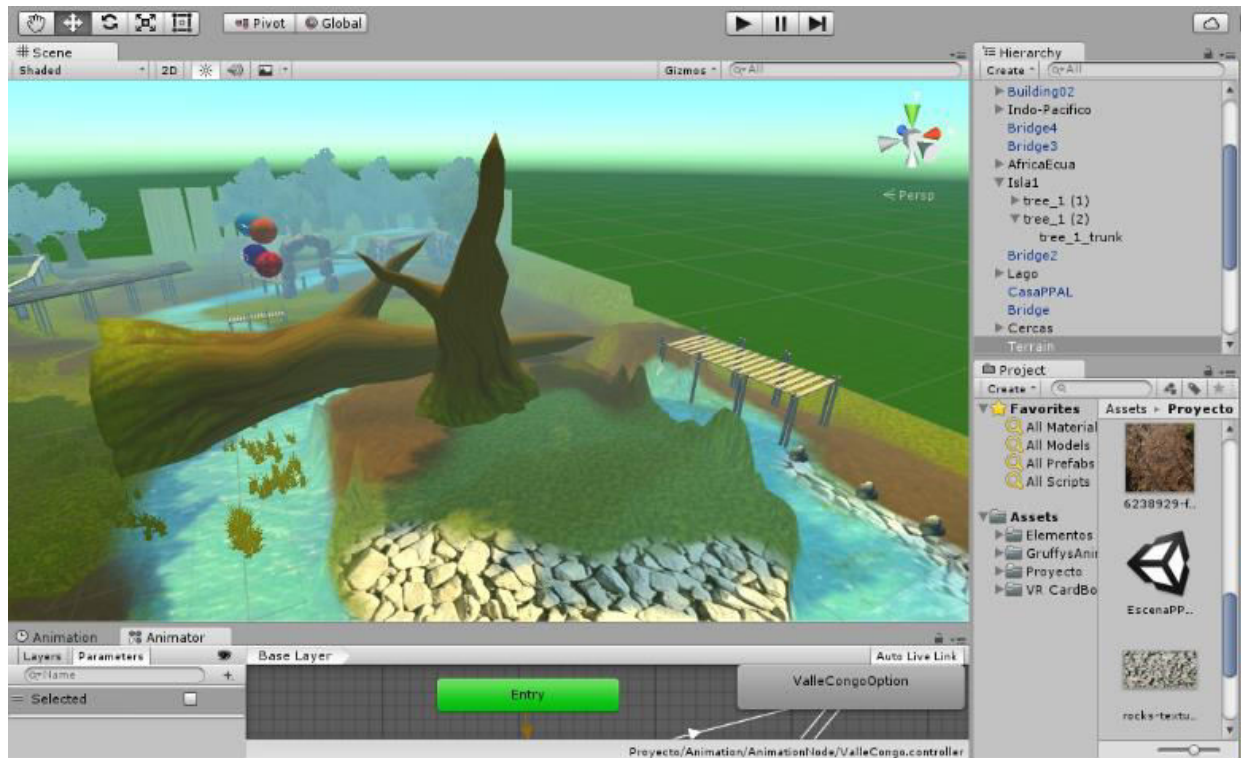


Figura 31. Representación de los humedales en el ambiente 3D que recrea la zona.

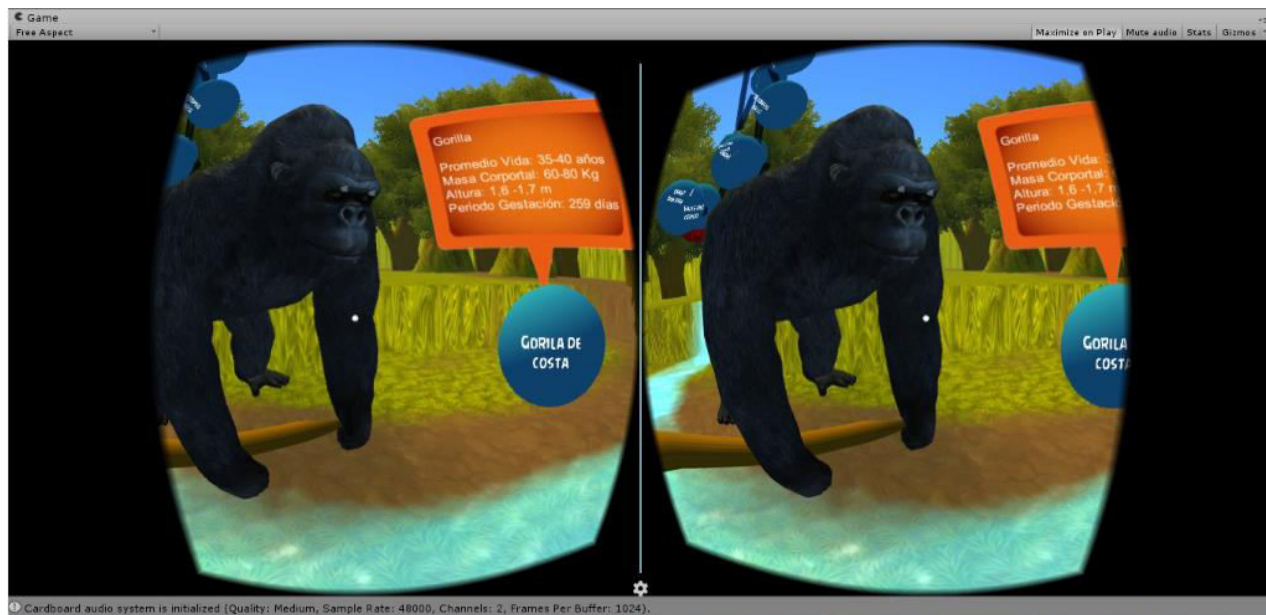


Figura 32. Representación de uno de los animales en el ambiente 3D.

El artículo “*User Experiences in Virtual Reality Environments Navigation Based on Simple Knowledge Organization Systems*” (Martin-Moncunill, Gaona-García, Gordillo-Orjuela, & Montenegro-Marin, 2016) muestra un primer análisis de la interfaz desarrollada, empleando el modelo TAM (“*Technology Acceptance Model*”) (Davis, Bagozzi, & Warshaw, 1989), que precisamente se escogió buscando obtener unas primeras impresiones sobre la introducción de este tipo de interfaces.

Los resultados del experimento animan a seguir trabajando en este tipo de interfaces, apuntando problemas a resolver como la necesidad de mejorar la navegación en horizontal o cierto mareo y sensación de estar perdido al navegar, especialmente cuando el KOS está completamente desplegado. Por otro lado, entre los problemas detectados se encontraron algunos como el mareo que han sido frecuentemente reportados en este tipo de tecnologías para entornos virtuales 3D (Treleaven, y otros, 2015).

El aspecto más relevante de estos dos experimentos en cuanto al objeto y alcance de esta tesis, lo constituye el hecho que para los dos experimentos realizados se trabajó con un KOS elaborado ad-hoc, dónde se tuvo en cuenta la usabilidad del propio KOS en relación a las funcionalidades que se recogerían en las interfaces. De esta manera se perseguía una mejora del entendimiento de la utilidad y la predisposición del usuario a emplear la interfaz.

Capítulo 5: Conclusiones

“Señores, estos son mis poderes.”

(Atribuida al Cardenal Cisneros ante la insolencia de los nobles castellanos que cuestionaron su legitimidad como regente, mientras señalaba tropas y cañones desde un balcón)

Las conclusiones particulares de las distintas investigaciones y experimentos recogidas en esta obra han sido plasmadas en los capítulos y secciones correspondientes. Este capítulo sirve como nexo de unión de los anteriores, dando cuenta de las conclusiones de una manera estratégica, relacionándolas claramente con los distintos objetivos de la tesis, las preguntas de investigación y las aportaciones realizadas.

El capítulo queda estructurado de la siguiente forma:

- En primer lugar, se retoman los objetivos y se indica su cumplimiento, empleando una tabla para cada aspecto tratado sobre los que se clasifican los objetivos, a saber:
 - o La catalogación automática de recursos empleando KOS de dominio específico.
 - o La evaluación de la especificidad de dominio en los términos que componen los KOS.
 - o La distancia de especialización entre conceptos.
- Posteriormente, se proporciona una respuesta concisa a las preguntas de investigación planteadas.
- Teniendo en cuenta la consecución de objetivos y la respuesta a las preguntas de investigación planteadas, se expondrá la aportación realizada.
- El capítulo cerrará con la conclusión general respecto al objeto principal del trabajo y futuras investigaciones.

La tabla 25 recuerda las preguntas de investigación formuladas y los objetivos establecidos, relacionándolas para un mejor entendimiento.

Tabla 25. Preguntas de investigación y objetivos asociados

Pregunta de investigación	Objetivos relacionados
¿Las interfaces basadas en KOS tienen potencial de uso según los estudios existentes?	Pregunta inicial que da origen a la tesis. La experimentación y análisis continuo de interfaces, permitió verificar el papel de la usabilidad del propio KOS, incluyendo los factores que se analizan en el capítulo 3.
¿Los algoritmos de catalogación automática de recursos constituyen una solución viable, proporcionando un nivel de calidad adecuado, frente al problema que puede constituir la catalogación manual?	Crear una metodología de evaluación de la calidad de las palabras clave extraídas automáticamente en contraste con las extraídas manualmente, que permita garantizar a los responsables de las colecciones de recursos una solución más adecuada al tradicional criterio de precisión/exhaustividad.
	Emplear un mecanismo de evaluación en un caso real y aportar los resultados para comprobar si la catalogación automática de recursos para el algoritmo analizado sería una solución viable.
¿Evaluar la especificidad de dominio en los términos que componen un KOS podría tener impacto en su uso?	Crear una metodología que permita evaluar la especificidad de dominio de un KOS, proporcionando indicadores para tal fin.
	Aplicar esta metodología para evaluar el grado de especificidad de dominio global de un KOS y clasificar los términos del KOS en función a su grado de especificidad de dominio y analizar el esfuerzo empleado en la aplicación de la metodología.
	Indagar sobre la posibilidad de emplear mecanismos automatizados para llevar a cabo la tarea.
¿Asignar un grado de distancia de especialización entre conceptos podría mejorar la usabilidad de un KOS?	Concretar un sistema que permita asignar valores de distancia de especialización entre conceptos y crear un KOS que incluya esa información para posibilitar su evaluación.
	Aplicar la metodología y obtener los resultados de distancia de especialización entre conceptos para un área de un KOS.
	Estudiar las posibilidades de uso de la distancia de especialización entre conceptos para una interfaz basada en texto y una interfaz visual.

5.1 Consecución de objetivos

Tabla 26. Consecución de los objetivos previstos.

a) Catalogación automática de recursos empleando KOS de dominio específico	
Garantizar a los responsables de las colecciones de recursos una solución para evaluar la calidad de las palabras clave extraídas automáticamente más adecuada al tradicional criterio de precisión/exhaustividad.	<p>Se ha propuesto una metodología de evaluación que compara el desempeño de humanos con el de algoritmos automáticos, más allá de la coincidencia con las palabras clave originales, que en experimentos anteriores se tomaba como patrón oro, aunque su calidad no hubiera sido verificada.</p> <p>El nuevo enfoque contrasta los resultados cuantitativos obtenidos mediante el criterio de precisión / exhaustividad, con un análisis cualitativo realizado por expertos donde los evaluadores desconocen la procedencia de las palabras clave: asignadas originalmente, asignadas por otro catalogador o asignadas por un algoritmo automático.</p> <p>Este tipo de evaluación proporcionaría a los responsables de las colecciones de recursos una mayor información respecto a la calidad de las palabras clave existentes en su colección y a los resultados que se obtendrían de emplear un algoritmo automático.</p>
Emplear este mecanismo de evaluación en un caso real y aportar los resultados para comprobar si la catalogación automática de recursos para el algoritmo analizado sería una solución viable.	<p>Se evaluó una selección de documentos de la colección de AGRIS en VOA3R, empleando el <i>framework</i> KEA. La muestra empleada para la comparación automática presentaba un tamaño y homogeneidad considerablemente más relevante que las empleadas en estudios anteriores; los resultados respecto a este punto fueron similares a los obtenidos en experimentos anteriores.</p> <p>La comparación del desempeño de expertos humanos con KEA al intentar hacer coincidir exactamente las palabras clave asignadas originalmente mostró que KEA "superó" a los seres humanos, incluso empleando un corpus de entrenamiento de tamaño limitado. Finalmente, en cuanto al análisis de la calidad de las palabras clave, los resultados obtenidos por KEA fueron similares a los obtenidos por humanos, lo que mostraba su aplicabilidad práctica como un servicio de indexación automática para el repositorio.</p>

b) Evaluación de la especificidad de dominio de los términos de un KOS	
Proporcionar un mecanismo de evaluación de la especificidad de dominio de un KOS.	<p>La evaluación propuesta se basa en un triple enfoque, que incluye análisis cualitativo realizado por expertos, cuantitativo y sistemático, contrastan los resultados para que complementen el análisis del estudio. La aplicación de esta metodología permite obtener indicadores en cuanto a:</p> <ul style="list-style-type: none"> - Grado de especificidad de dominio para cada término del KOS mediante un valor numérico que se emplea para su clasificación como “Genérico”, “Incierto” o “Específico”. - Grado de especificidad de dominio para cada una de las ramas del KOS analizado, proporcionando un método de cálculo basado en pesos e inspirado en el criterio de precisión / exhaustividad. <p>Adicionalmente, la metodología solicita a los evaluadores proporcionar información cualitativa adicional respecto a sus decisiones o problemas a la hora de realizar su evaluación.</p>
Aplicar este mecanismo para evaluar el grado de especificidad de dominio global de un KOS y clasificar los términos del KOS en función a su grado de especificidad de dominio y analizar el esfuerzo empleado.	<p>La metodología se aplicó sobre 4 ramas del tesoro AGROVOC. La evaluación manual del grado de especificidad de dominio requirió una elevada cantidad de tiempo y resultó ser más complejo de lo esperado, dado que los términos están sujetos a diferentes interpretaciones y opiniones subjetivas que llevan a los expertos a desacuerdos o simplemente no les permiten ser claramente clasificados como de dominio específico o genéricos. Con el fin de obtener mayor información sobre el razonamiento de los expertos detrás de la asignación del "grado de especificidad" de un término se realizó un análisis cualitativo de los casos inciertos.</p>
Indagar sobre la posibilidad de emplear mecanismos automatizados para facilitar la evaluación de especificidad de dominio de un KOS y categorizar sus términos.	<p>El acercamiento consistió en contrastar los resultados obtenidos con las ontologías SUMO y MILO suponiendo que, al ser estas ontologías no específicas, los términos considerados "específicos" por expertos humanos no deberían existir en las mismas. La cantidad de términos encontrados en SUMO y MILO muestra que no es posible una utilización directa, pero los resultados validan los supuestos iniciales en el sentido que el número de términos no específicos encontrados fue significativamente mayor que el de términos inciertos y mucho más alto en comparación con el número de términos específicos encontrados. Esto también se refleja en el hecho de que los términos encontrados en MILO fueron más que en SUMO y con una distribución similar.</p>

c) Posibilidades de uso de la distancia de especialización entre conceptos, tanto en interfaces visuales como basadas en texto	
Proponer un sistema que permita asignar valores de distancia de especialización entre conceptos a KOS.	Dado que no se encontró ningún KOS ni área de KOS que incluyera información sobre la distancia de especialización entre conceptos, se propuso un método en el que los valores eran asignados por humanos, usuarios potenciales del KOS a analizar, dónde estos evaluaban subjetivamente, empleando una escala de cinco puntos y posteriormente se contrastaban los resultados estadísticamente para fijar los valores de las distancias definitivos.
Aplicar este sistema para obtener un KOS que incluya información sobre la distancia de especialización entre sus conceptos.	<p>Se obtuvieron resultados de 21 participantes, los cuales mostraron un nivel de acuerdo que permitió la elaboración del set de datos de distancias de especialización entre conceptos, en base a los cuales se fijaron las distancias definitivas para la elaboración de las interfaces. Además, la metodología empleada para la asignación de la distancia de especialización resultó ser útil para detectar inconsistencias en el KOS y podría ayudar en el control de calidad y optimización de las relaciones jerárquicas.</p> <p>Puesto que el acercamiento para fijar las distancias, mediante una escala de 5 puntos, no había sido empleado con anterioridad, los participantes fueron preguntados sobre la idoneidad del empleo de esta escala. Los 21 participantes declararon que consideraban adecuado usar una escala de cinco puntos para la selección de términos. Más de la mitad subrayaron que en el caso de utilizar una escala mayor habrían tenido problemas para asignar valores intermedios.</p>
Emplear el KOS obtenido para realizar una primera evaluación sobre las posibilidades de uso de la distancia de especialización entre conceptos en interfaces basadas en texto e interfaces visuales.	Se elaboraron los dos tipos de interfaces, las cuales fueron evaluados siguiendo un enfoque de prueba A / B en el que se enfrentaban los casos que consideraban la distancia de especialización y los que no lo hacían. Se presentaron cinco casos de búsqueda a 40 participantes que no participaron en la parte anterior del experimento relativa a la asignación de la distancia de especialización entre conceptos. Los resultados obtenidos mostraron un gran potencial para la interfaz basada en texto, mientras que las expectativas para la interfaz visual permanecieron más bajas, aunque la mayoría de los participantes encontraron que el enfoque podría resultar útil tanto para interfaces visuales como basadas en texto una vez que se les explicó la lógica de la construcción de la interfaz.

--	--

5.2 Respuesta a las preguntas de investigación

1. ¿Las interfaces basadas en KOS tienen potencial de uso según los estudios existentes?

<p>El análisis del estado de la cuestión, contrastado con los estudios realizados por el autor en el área <u>muestran claramente el potencial de uso de las interfaces que emplean KOS.</u></p>

Tal como se comentó anteriormente, los estudios de interfaces fueron compilados en un análisis de la literatura realizado con un enfoque sistemático. A través de estos estudios se identificaron distintos problemas de usabilidad en las interfaces y recomendaciones para la elaboración de las mismas. Respecto a estas recomendaciones habría que tener en cuenta que cada interfaz debe responder a usuarios específicos con objetivos en un contexto de uso específico, por lo que cualquier recomendación heurística debería ser revisada conforme a estos tres puntos antes de asumir su validez para el caso concreto.

La recomendación general sería basarse en los principios del diseño centrado en el usuario para el desarrollo de estas interfaces; sin embargo, su desarrollo parte de unas restricciones impuestas por las particularidades del KOS a emplear, lo que a juicio del autor constituye el problema principal respecto al correcto desarrollo, adopción y uso de las interfaces que permiten la visualización de los esquemas generados a partir de los términos y relaciones de los KOS.

Estos problemas, que están por encima del diseño de la interfaz y ligados a la usabilidad del KOS, destacan dificultades en cuanto a la percepción de la utilidad del KOS, el entendimiento del mismo y la necesidad de aprender sus particularidades, lo que limita en gran medida su adopción y uso. La identificación de estos problemas motivó el análisis de los distintos factores recogidos en esta tesis, a saber:

- La catalogación automática de recursos empleando KOS de dominio específico.
- La especificidad de dominio en los términos que componen un KOS.
- La distancia de especialización entre conceptos de un KOS.

2. *¿Los algoritmos de catalogación automática de recursos constituyen una solución viable, proporcionando un nivel de calidad adecuado, frente al problema que puede constituir la catalogación manual?*

El estudio realizado en la colección de VOA3R-AGRIS evidencia que el algoritmo empleado, no solamente muestra un nivel de desempeño superior al de los expertos humanos al intentar hacer coincidir exactamente las palabras clave asignadas originalmente, sino que en la evaluación cualitativa sus resultados fueron similares a los obtenidos por humanos, lo cual indica su aplicabilidad práctica como un servicio de indexación o catalogación automática.

El análisis de la literatura respecto a los algoritmos de extracción de palabras clave existentes muestra que estos proporcionan resultados positivos al contrastar las palabras clave seleccionadas con las asignadas originalmente. Sin embargo, este tipo de evaluación dejaba de lado otros aspectos relacionados con la calidad de las palabras clave extraídas y daba por supuesto que las asignadas originalmente eran las que mejor se adecuaban.

En el estudio planteado, adicionalmente a la comparación en términos de precisión exhaustividad que comparaba las palabras clave originales con las asignadas por el algoritmo automático, se comparó el índice de acierto entre expertos catalogadores humanos tratando de realizar la misma tarea y finalmente se solicitó a estos expertos que analizaran la calidad de las palabras clave asignadas, sin que fueran conscientes de su procedencia.

De este experimento se concluyó que los resultados del algoritmo automático KEA son comparables a los de los humanos, no solo desde un punto de vista cuantitativo (coincidencias exactas de palabras clave con las originalmente asignadas), sino también cualitativo; en base al nuevo enfoque de evaluación propuesto.

Dado el reducido coste y la facilidad de aplicación de este tipo de mecanismos, la necesidad de catalogación de recursos como problema relativo a la adopción y uso de KOS, quedaría resuelta en gran medida por los mismos. Además, estos mecanismos podrían resultar de ayuda como herramienta de apoyo a la catalogación de recursos, sugiriendo y revisando automáticamente términos a los catalogadores mientras estos van cumplimentando los metadatos correspondientes.

Así, para el caso del algoritmo KEA y teniendo en cuenta los resultados de estudios anteriores, se podría concluir que constituye una solución viable al problema de catalogación de recursos. En lo que respecta a otros algoritmos, se podría aplicar la metodología propuesta para obtener evidencias adicionales.

3. *¿Evaluar la especificidad de dominio en los términos que componen un KOS podría tener impacto en su uso?*

La existencia de términos específicos de dominio dota a los KOS de una utilidad que puede quedar limitada cuando estos conviven con conjuntos de términos no específicos de dominio, lo cual es una situación frecuente. Identificar estos términos puede mejorar la utilidad del KOS y por tanto extender su uso. Adicionalmente, el listado de términos específicos de dominio extraídos tendría utilidad por sí mismo.

Respecto al estudio realizado con AGROVOC, se identificaron casos característicos en los que la identificación de la especificidad de los términos sería de ayuda:

- a) Emplear AGROVOC para identificar la temática de recursos sin clasificar; los términos no específicos de dominio existentes podrían llevar a una catalogación errónea.
- b) Los repositorios o librerías digitales que sólo deseen trabajar con los términos específicos del dominio de agricultura en AGROVOC no podrían adoptar directamente AGROVOC tal como se encuentra el KOS actualmente, dado el número de términos no específicos que incluye.
- c) Sistemas inteligentes que empleen el KOS de dominio específico AGROVOC para intentar deducir el dominio de un recurso podrían presentar problemas, dada la cantidad de términos de otros dominios (por ejemplo, economía) que este incluye.

Se ha propuesto una metodología para la identificación de estos términos que además ha mostrado potencial para detectar inconsistencias en los KOS y mejorar su usabilidad. Esta metodología permite establecer valores para medir el grado de especificidad de dominio para cada término del KOS mediante un valor numérico que se emplea para su clasificación como “Genérico”, “Incierto” o “Específico”; así como el grado de especificidad de dominio para cada una de las ramas del KOS analizado, proporcionando un método de cálculo basado en pesos e inspirado en el criterio de precisión / exhaustividad.

La aplicación de la metodología requiere de una elevada cantidad de tiempo y puede resultar compleja. No existe un acercamiento inmediato para obtener la especificidad de un término de forma automática, pero el contraste de los resultados empleando las ontologías SUMO y MILO fue de utilidad para validar los supuestos iniciales.

4. *¿Asignar un grado de distancia de especialización entre conceptos podría mejorar la usabilidad de un KOS?*

Los resultados del estudio realizado sobre un KOS construido incluyendo información sobre la distancia de especialización entre conceptos muestran que asignar un grado de distancia de especialización entre conceptos ha mejorado la usabilidad del mismo, en su aplicación a una interfaz visual y otra basada en texto; en comparación con el uso de las mismas interfaces sin emplear esa información.

El entendimiento de las relaciones jerárquicas tipo "gen-spec" como "todo o nada", asumiendo que la relación siempre es igualmente fuerte entre un clasificador y cualquiera de sus clasificadores relacionados mediante una relación "gen-spec" y también en todos los niveles de la jerarquía. Este hecho que, según diversos autores, puede constituir una simplificación excesiva del relato psicológico de las relaciones del mundo real, puede llegar a dificultar el entendimiento y uso del KOS.

Este hecho quedó patente en el experimento realizado sobre las posibilidades de uso de la distancia de especialización entre conceptos, pero también en el experimento anterior en cuanto a la evaluación de la especificidad de dominio; donde los usuarios llegaron en algunos casos a afirmar que el KOS presentaba inconsistencias.

La asignación de un grado de distancia relativa para representar el nivel de similitud entre los pares de clasificadores relacionados dotaría al KOS de una información de utilidad, que podría ser empleada para establecer pesos que se tengan en cuenta para mejorar los mecanismos de sugerencia de términos de búsqueda o resultados relacionados en interfaces textuales y adicionalmente para decidir sobre la representación de los diferentes clasificadores y su posición en la pantalla en interfaces visuales.

Los resultados obtenidos mostraron un gran potencial para la interfaz basada en texto. Las expectativas para la interfaz visual fueron más bajas, probablemente debido a que la implementación de esta interfaz no resultó ser la más adecuada. En cualquier caso, la mayoría de los participantes encontraron que el enfoque podría resultar útil tanto para interfaces visuales como basadas en texto una vez que se les explicó la lógica de la construcción de la interfaz y lo que implicaba la distancia de especialización entre conceptos.

Finalmente, la metodología empleada para la asignación de la distancia de especialización resultó ser útil para detectar inconsistencias en el KOS y podría ayudar en el control de calidad y optimización de las relaciones jerárquicas.

5.3 Aportaciones

Las aportaciones principales de la investigación realizada están relacionadas con:

- La creación de metodologías para la evaluación de aspectos ligados con la adopción y uso de los KOS, concretamente referente a:
 - o La catalogación de recursos basada en KOS específicos de dominio mediante algoritmos automáticos, buscando un sistema de evaluación que proporcionara más información sobre su calidad comparada al trabajo que podrían realizar catalogadores humanos.
 - o El nivel de especificidad de dominio de los términos de un KOS.
 - o La distancia de especialización entre conceptos, evaluando la jerarquía de un KOS.

Los anexos II, III y IV presentan una guía de aplicación para cada una de estas metodologías

- Los resultados obtenidos de la aplicación de esas metodologías.

También son resultados destacables:

- Los sets de datos obtenidos para poder proceder a la realización de las evaluaciones. Estos sets de datos podrían ser empleados para realizar experimentación adicional o para entrenar futuros mecanismos automáticos.
- La labor realizada en el análisis del estado del arte y el llevado a cabo en distintas interfaces visuales.

En lo referente a la labor realizada en el análisis de interfaces y visualización de KOS, se debe tener en cuenta que este tema no era el objeto principal del estudio y no se ha profundizado sobre el mismo. En cualquier caso, aunque los resultados y aportaciones en esta área sólo han sido resumidos en los capítulos anteriores de la presente tesis; toda la investigación llevada a cabo hasta el momento ha sido publicada en revistas de investigación y es accesible a través de las mismas.

Las siguientes subsecciones proporcionan información adicional sobre las aportaciones realizadas, dividiendo las mismas en función de las áreas en las que se enmarcaban los objetivos de la tesis.

5.3.1 Aportaciones respecto a la catalogación automática de recursos empleando KOS de dominio específico.

La catalogación automática de recursos empleando algoritmos que hacen uso de KOS para identificar y extraer palabras clave puede resultar un mecanismo que ahorre muchos costes, tanto en tiempo como en esfuerzo, respecto a la catalogación manual.

El tipo de evaluación utilizado hasta la fecha de la realización del estudio empleaba medidas de precisión y exhaustividad comparando la lista de palabras clave generadas a las listas proporcionadas por el autor o catalogador original. Esto es, las palabras clave asignadas originalmente se tomaban como el patrón válido respecto al cual debía juzgarse la calidad de los resultados del algoritmo.

Este enfoque no asegura la aplicabilidad práctica de las técnicas, en el sentido que puede plantearse el responsable de un repositorio, cuyo objetivo será encontrar un algoritmo que pueda obtener una calidad similar a la de un catalogador humano.

Un enfoque basado únicamente en la coincidencia exacta de términos parte de la base que las palabras clave elegidas originalmente eran las más apropiadas y no responde a la pregunta de si otros catalogadores humanos hubieran obtenido mejores resultados al tratar de coincidir con las palabras clave asignadas por el autor original, ni refleja cuál sería la calidad relativa de las palabras clave asignadas por el ser humano en comparación con las asignadas de forma automática.

La metodología aplicada en el estudio plasmado en esta tesis sí permite contestar a estas preguntas, pues emplea tres criterios para evaluar la calidad de las palabras clave extraídas:

- 1) El enfoque tradicional de coincidencia exacta del algoritmo con las palabras clave originales, medido en términos de precisión y exhaustividad.
- 2) La coincidencia exacta de catalogadores humanos con las palabras clave originales; enfrentando estos resultados a los conseguidos por el algoritmo.
- 3) La evaluación subjetiva de la calidad de las palabras clave asignadas, por parte de catalogadores humanos, sin especificar cómo se habían extraído.

Los resultados obtenidos aportaron evidencias adicionales sobre la efectividad de KEA, puesto que coincidían con los que se recogen en experimentos anteriores. Adicionalmente, el experimento se realizó con una colección de tamaño considerablemente mayor y de temática más heterogénea que los existentes hasta la fecha.

De la evaluación realizada se pudo concluir que el uso del algoritmo KEA con AGROVOC para la catalogación automática de recursos en el dominio que trata este KOS resultaría una solución adecuada a la catalogación manual. También podría ser empleado como una herramienta de apoyo a la catalogación manual, sugiriendo términos, pues tanto KOS como AGROVOC, que se mueven en magnitudes de decenas de miles de términos, resultaría de gran ayuda.

5.3.2 Aportaciones respecto a la evaluación de la especificidad de dominio en los términos que componen los KOS.

El origen de un KOS, en la mayoría de los casos, está motivado por la necesidad de organizar la información en un dominio completo y así aplicarlo a la resolución de problemas dentro del mismo. De esta forma, se espera que dentro de un KOS específico de dominio se encuentren términos que sean claramente específicos de ese dominio.

Sin embargo, según estos KOS van adquiriendo tamaño o por necesidades específicas del contexto en el que se les va a dar uso, es frecuente que contengan una gran cantidad de términos generalistas o específicos de otras áreas.

Como ejemplo de esto, en el estudio llevado a cabo en AGROVOC, se pudieron encontrar un buen número de términos relacionados con la economía. Sin lugar a dudas, estos términos económicos se aplican en la agricultura y pueden ser necesarios en el contexto de uso de AGROVOC, pero no son términos representativos del dominio en sí mismos, esto es, si aislamos ese conjunto de términos no formaríamos un KOS específico del dominio de la agricultura, sino del de la economía.

Las listas de términos específicos de un dominio permiten identificar y representar ese momento, teniendo un gran valor por sí mismas, sin siquiera estar formando parte de un KOS que defina las relaciones entre los términos. Es por ello que el desarrollo de una metodología que permita catalogar la especificidad de los términos específicos de dominio contenidos en un KOS, podría dotar de una mayor utilidad y usabilidad al mismo.

El trabajo realizado aporta una metodología que permite identificar los términos específicos de dominio y establecer una puntuación de “especificidad de dominio” al KOS analizado, basándose en 3 pilares:

1. Un análisis cualitativo en el que expertos en el dominio asignan valores a los términos, basados en los principios sobre teoría de dominio establecidos por los autores más reconocidos en el área (Hjørland, Albrechtsen y Mai).

2. Un análisis cuantitativo estableciendo parámetros que analizan los resultados obtenidos, procedentes del análisis de expertos que permite calcular:
 - a. El grado de especificidad de dominio para cada término del KOS mediante un valor numérico que se emplea para su clasificación como “Genérico”, “Incierto” o “Específico”.
 - b. El grado de especificidad de dominio para cada una de las ramas del KOS analizado, proporcionando un método de cálculo basado en pesos e inspirado en el criterio de precisión / exhaustividad.
3. Un enfoque sistemático que permitiría su aplicación en otro tipo de KOS, modificando el sistema de pesos tal como se comenta en el capítulo que desarrolla el experimento.

La metodología se aplicó con éxito a 4 ramas representativas del tesoro AGROVOC. Dado que durante el desarrollo del experimento se evidenció la necesidad de invertir una considerable cantidad de tiempo y contar con expertos del dominio para poder evaluar la especificidad de una lista de términos; se consideró la posibilidad de contar en un futuro con mecanismos automáticos que llevaran a cabo estas tareas. Como primer acercamiento, se contrastaron los resultados obtenidos por los expertos humanos con las ontologías generalistas SUMO y MILO, entendiendo que, al no ser estas ontologías específicas, los términos considerados "específicos" por expertos humanos no deberían existir en las mismas.

La cantidad de términos encontrados en SUMO y MILO no permite el uso de estas ontologías directamente para obtener un mecanismo de catalogación automático, pero resultó de utilidad para validar los supuestos iniciales en el sentido que el número de términos no específicos encontrados fue significativamente mayor que el de términos inciertos y mucho más alto en comparación con el número de términos específicos encontrados. Además, los términos encontrados en MILO fueron más que en SUMO y con una distribución similar.

El set de datos resultante de la evaluación de los términos, es una aportación valiosa en sí misma, puesto que:

- 1) Constituye una lista de términos específicos del dominio que trata AGROVOC.
- 2) Puede ser empleada para entrenar a sistemas inteligentes a discernir en cuanto a la especificidad de dominio.
- 3) Supone un primer paso que permitiría hacer futuras validaciones del experimento o ser de utilidad para experimentos similares, dado que el set de datos incluye las puntuaciones y la catalogación de términos específicos, no específicos y no claros.

Por último, la aplicación de una metodología de este tipo permitiría realizar una revisión del KOS de la que extraer mejoras adicionales, aunque no estuvieran directamente ligadas a la especificidad de dominio de los términos. De hecho, los sujetos que participaron en el experimento plantearon incoherencias relativas a la jerarquía del tesoro, que motivaron aún más la realización del experimento relativo a la distancia de especialización entre conceptos.

5.3.3 Aportaciones respecto a las posibilidades de uso de la distancia de especialización entre conceptos, tanto en interfaces visuales como basadas en texto

El entendimiento común de las relaciones del tipo generalización / especialización ha sido considerado por varios autores como una simplificación excesiva del relato psicológico de las relaciones del mundo real, dado que son consideradas como "todo o nada", asumiendo que la relación siempre es igualmente fuerte entre un clasificador y cualquiera de sus clasificadores relacionados mediante una relación de generalización / especialización y también en todos los niveles de la jerarquía.

Ante esto se ha propuesto la asignación de un grado de distancia relativa para representar el nivel de similitud entre los pares de clasificadores relacionados, denominada "distancia de especialización entre conceptos"; considerando que esta podría ser valiosa para fines de búsqueda y se podría aplicar a la representación visual de los KOS, facilitando además su entendimiento.

Sin embargo, más allá de un número reducido de propuestas teóricas, no se han encontrado casos que hayan evaluado la aplicación de esta distancia de especialización entre conceptos; más aún, no se ha encontrado ningún KOS que incluya esta información para poder evaluar la utilidad de contar con ella.

Así, la primera aportación del estudio realizado en cuanto a las posibilidades de uso de la distancia de especialización entre conceptos, sería la propia elaboración del set de datos de distancias de especialización entre conceptos, resultando:

- 1) El método empleado para la elaboración del set de datos de distancia de especialización entre conceptos.
- 2) El set en sí mismo, que permitiría la realización de experimentos adicionales sobre el uso de la distancia de especialización entre conceptos.

Adicionalmente, el método empleado para la elaboración del set de datos de distancia de especialización entre conceptos resultó de utilidad para detectar inconsistencias en el KOS y podría ser aplicado para tal fin.

En cuanto a la evaluación de las posibilidades de uso de la distancia de especialización entre conceptos, se siguió un enfoque de prueba A / B proponiendo casos de uso implementados a través de interfaces tanto visuales como basadas en texto, donde 40 participantes (usuarios potenciales), que no participaron en la parte anterior del experimento relativa a la asignación de la distancia de especialización entre conceptos; evaluaron 5 casos de usos para cada tipo de interfaz (basada en texto y visual).

La metodología de evaluación (basada en pruebas tipo A/B y entrevistas) no supuso ninguna aportación en sí misma, puesto que las pruebas del tipo A/B son sobradamente conocidas y su utilidad está sólidamente identificada y comprobada. La aportación reside por tanto en los resultados de la evaluación, que sería la primera realizada hasta la fecha, en cuanto a la aplicabilidad práctica del uso de información sobre la distancia de especialización entre conceptos.

Estos primeros resultados mostraron un gran potencial para la interfaz basada en texto, mientras que las expectativas para la interfaz visual permanecieron más bajas. El potencial menor apreciado para la interfaz visual podría deberse a que el enfoque de visualización empleado para el experimento no era la opción más adecuada para mostrar plenamente el potencial de la integración de la información, tal como reflejaron algunos participantes durante las entrevistas. Por otro lado, una vez que se les explicó la lógica de la construcción de la interfaz, la mayoría de los participantes encontraron que el enfoque podría resultar útil tanto para interfaces visuales como basadas en texto.

5.4 Trabajo futuro y conclusión final

El objeto de estudio del presente trabajo residía en indagar sobre las posibles causas de desuso de interfaces que permitían interactuar con KOS, en contraposición al potencial teórico reflejado en multitud de estudios existentes y contribuir aportando metodologías que permitieran evaluar su impacto, proporcionando así información de utilidad para modificar un KOS de forma que se potencie su uso.

Las causas identificadas que se tratan en esta obra parten del trabajo realizado en la evaluación de las interfaces presentes en el portal Organic.Edunet y fueron igualmente identificadas en estudios de interfaces posteriores realizados por el autor, todos ellos actualmente publicados en distintas revistas y actas de congresos. El estudio sobre uso de mecanismos de catalogación automática, pareció resultar el que debiera realizarse con mayor premura, puesto que, si no es posible catalogar adecuadamente los recursos en base a un KOS, no tendría sentido emplear una interfaz que haga uso de ese KOS.

Este estudio mostró que el problema que supone la necesidad de catalogar los recursos con KOS es fácilmente superable. Los estudios sobre especificidad de dominio y distancia de especialización entre conceptos mostraron la posibilidad de mejorar la utilidad y la usabilidad de los KOS; pero la no existencia de este tipo de mecanismos en los KOS actuales no se explica por sí misma, ni se podría asegurar que sea de tal importancia como para estar impidiendo la adopción y uso de KOS en relación a los estudios existentes.

Sin embargo, la experiencia obtenida en estos estudios junto con la extraída de los realizados en el ámbito del análisis de interfaces visuales que permiten interactuar con KOS, sí señala claramente que la usabilidad del propio KOS es un aspecto escasamente considerado en el diseño de interfaces. Durante el desarrollo de este tipo de interfaces muy raramente se tienen en cuenta las características propias y problemas de usabilidad del KOS que será empleada por usuarios específicos, con objetivos específicos, en un contexto de uso específico.

El hecho de que los KOS se elaboren (o se parta de algunos ya elaborados) de forma completamente ajena a su uso en interfaces, supone que la integración de los KOS con las interfaces que tienen que hacer uso de los mismos, se convierta en un auténtico quebradero de cabeza e imposibilite la construcción de interfaces que satisfagan correctamente las necesidades del usuario.

Por otro lado, esto no respondería completamente al motivo por el que, en los estudios realizados a nivel teórico, el uso de estas interfaces, aun no habiendo tenido en cuenta la usabilidad del KOS, se obtengan resultados positivos. Para comprender la explicación que en opinión del autor responde a este fenómeno, conviene dejar primeramente clara la diferencia entre utilidad, usabilidad, deseo de uso y experiencia de usuario. La figura 33 muestra un esquema que facilita la comprensión de estos conceptos.

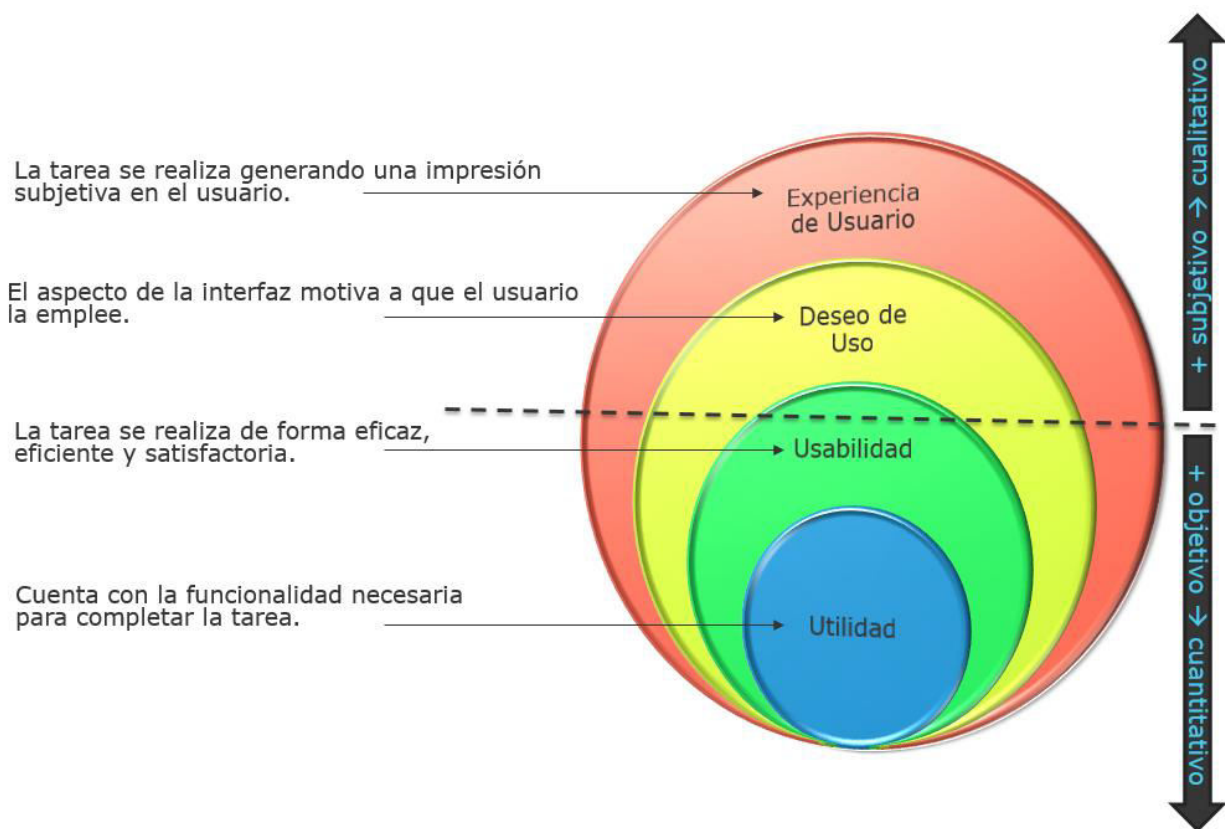


Figura 33. Utilidad, usabilidad, deseo de uso y experiencia de usuario.

La utilidad se refiere a la capacidad de un sistema para que el usuario pueda mediante el mismo completar la tarea asignada. Una vez el sistema dispone de esta utilidad, entra en juego la usabilidad que se enmarca en el cómo; esto es, la manera en la que se realiza esa tarea que debería ser efectiva, eficiente y satisfactoria para el usuario.

La utilidad y la usabilidad pueden ser medidas mayoritariamente y con facilidad desde un enfoque cuantitativo y centrándose en hechos objetivos, pero los hechos subjetivos cuentan también con un papel fundamental. Estos hechos subjetivos los forman el deseo de uso y la experiencia de uso, los cuales deben ser valorados con un enfoque esencialmente cualitativo.

El deseo de uso está relacionado con las características del sistema que motivan al usuario a emplearlo y que están directamente relacionadas con el “*look & feel*” de la interfaz, mientras que la experiencia de usuario lo está con los sentimientos y las sensaciones que tiene el usuario mientras hace uso del sistema a través de la interfaz.

Teniendo esto en cuenta, resulta claro que, si un usuario no es capaz siquiera de apreciar la utilidad de un sistema, no va a sentir ninguna predisposición a usarlo y menos aún si esto le va a suponer un esfuerzo para poder aprender a usar una interfaz o comprender la clasificación de un KOS.

El motivo de porqué los experimentos realizados hasta la fecha puedan estar arrojando resultados prometedores a nivel teórico mientras que a nivel práctico la situación no varía, podría tener en realidad una explicación bastante simple: en cualquier tipo de test existen unos usuarios a los que se les ha pedido que prueben algo, por lo que estos inevitablemente lo van a probar, se esforzarán en entender su funcionamiento y dar su opinión respecto; en el mundo real si un usuario no es capaz de percibir la utilidad de algo, en la mayoría de los casos ni siquiera llegará a usarlo.

Suponiendo que llegara a usarlo, este usuario debería realizar un esfuerzo por comprender el esquema de clasificación del KOS e invertir un tiempo en aprender sobre el mismo y sobre el propio uso de la interfaz. Si el usuario no tiene ningún tipo de motivación para invertir tiempo en este aprendizaje (porque desconoce la utilidad de hacerlo) o si el KOS resulta incomprensible por haber sido integrado de forma inadecuada, irremisiblemente el usuario no empleará la interfaz.

Así, los futuros estudios sobre interfaces visuales que hagan uso de KOS deberían tener en cuenta todo lo expuesto anteriormente o de lo contrario los resultados correrán el riesgo de no pasar el ámbito teórico y estarán probablemente sesgados.

Concretamente el esfuerzo futuro debería centrarse en:

- i. Facilitar el empleo y mejorar los mecanismos de catalogación automática con KOS específicos de dominio para aliviar la carga que supone esta necesaria tarea. En repositorios donde la participación de usuarios sea efectiva, deberían considerarse mecanismos de catalogación colaborativa que podrían convivir con los automáticos y los realizados por expertos.
- ii. La creación y mejora de mecanismos que permitan evaluar aspectos ligados con la usabilidad y la utilidad de los KOS y permitan mejorarlo, como los detectados en el marco del presente trabajo respecto a la especificidad de dominio y la distancia de especialización entre conceptos. Estas recomendaciones deberían tenerse en cuenta desde el momento de creación de los KOS, incluso podrían incorporarse como parte de las metodologías existentes para la creación y mantenimiento de KOS, puesto que su aplicación resultaría mucho menos costosa que para KOS establecidos como AGROVOC.

- iii. La necesidad de adaptar o crear KOS teniendo en cuenta la función que van a realizar en una interfaz empleada por usuarios específicos con objetivos específicos en un contexto de uso específico.
- iv. Mostrar claramente al usuario final la utilidad de las interfaces que hagan uso de KOS y aportar un valor diferenciador sobre otro tipo de interfaces que posibiliten realizar las mismas tareas. Los casos de uso deben identificarse con gran precisión.
- v. El trabajo futuro específico para cada uno de los factores que podrían estar afectando a la adopción y uso de los KOS estudiados en esta obra ya fueron expuestos en las secciones correspondientes del capítulo 3.

La investigación respecto a la mejora de la utilidad y usabilidad de los KOS es un campo en el que cabe profundizar y especialmente en lo relativo a las posibilidades respecto a su uso en contextos de aprendizaje. Incorporar los conocimientos adquiridos para la mejora de las metodologías existentes que abordan la creación y mantenimiento de KOS y continuar progresando en el desarrollo y evaluación de interfaces que permitan interactuar con KOS, constituye la línea principal de trabajo futuro.

Existen multitud de metodologías para la construcción de KOS (Iqbal, Murad, Mustapha, & Sharef, 2013), algunas de ellas con enfoques muy novedosos (Zhitomirsky-Geffet, Erez, & Judit, 2017); que podrían ser optimizadas si se enmarcaran en un modelo de desarrollo que siguiera los principios del diseño centrado en el usuario, empleando técnicas de evaluación como las plasmadas en el presente trabajo. La definición de un marco de este tipo, que fuera amoldable a cualquier metodología de desarrollo de KOS supondría una importante aportación.

Referencias Bibliográficas

- Aitchison, J., Gilchrist, A., & Bawden, D. (2000). *Thesaurus construction and use: a practical manual*. Londres, Inglaterra: Aslib IMI, Stone House Court.
- Al-Hashemi, R. (2010). Text Summarization Extraction System (TSES) Using Extracted Keywords. *International Arab Journal of e-Technologies*, 1(4), 164-168.
- Anibaldi, S., Jaques, Y., Celli, F., Stellato, A., & Keizer, J. (2013). Migrating bibliographic datasets to the Semantic Web: The AGRIS case. *Proceedings of Semantic Web EFITA conference*. Torino, Italia.
- ANSI/NISO. (2005). *Z39.19-2005 (R2010) Guidelines for the Construction, Format, and Management of Monolingual Controlled Vocabularies*. Obtenido de http://www.niso.org/apps/group_public/project/details.php?project_id=46
- Aula, A., & Käkki, M. (2005). Less is more in Web search interfaces for older adults. *First Monday*, 10(7).
- Bates, M. J. (1989). The design of browsing and berrypicking techniques for the online search interface. *Online Information Review*, 13(5), 407-424.
- Bevan, N. (1999). Quality in use: Meeting user needs for quality. *Journal of systems and software*, 49(1), 89-96.
- Broughton, V. (2006). The need for a faceted classification as the basis of all methods of information retrieval. *Aslib Proceedings*, 58(1/2), 49-72.
- Burns, M., & Schell, G. P. (2002). Merlot: a repository of e-learning objects for higher education. *E-service Journal*, 1(2), 53-64.
- Buttenfield, B. (1999). Usability evaluation of digital libraries. *Science & Technology Libraries*, 17(3 & 4), 39-59.
- Cabré, M. T. (2005). *La terminología: representación y comunicación: elementos para una teoría de base comunicativa y otros artículos*. Barcelona, España: Documenta Universitaria. Institut Universitari de Lingüística Aplicada, Universidad de Barcelona.
- Cao, N., Sun, J., Lin, Y. R., Gotz, D., Liu, S., & Qu, H. (2010). Facetatlas: Multifaceted visualization for rich text corpora. Visualization and Computer Graphics. *IEEE Transactions on visualization and computer graphics*, 16(6), 1172-1181.
- Card, S., Mackinlay, J., & Shneiderman, B. (1999). *Readings in information visualization: using vision to think*. San Francisco, CA, USA: Morgan Kaufmann Publishers .
- Cechinel, C., Sánchez-Alonso, S., & Sicilia, M. A. (2009). Empirical analysis of errors on human-generated learning objects metadata. En *Metadata and semantic research* (págs. 60-70). Springer.

- Cerri, R., Barros, R. C., Freitas, A. A., & de Carvalho, A. C. (2014). Evolving relational hierarchical classification rules for predicting gene ontology-based protein functions. . *Proceedings of the 2014 conference companion on Genetic and evolutionary computation companion* (págs. 1279-1286). Philadelphia, PA, USA: ACM.
- Chandrasekaran, B., Josephson, J. R., & Benjamins, V. R. (1999). What are ontologies, and why do we need them? *IEEE Intelligent Systems and their applications*, 14(1), 20-26.
- Chen, S., & Chen, H. (2012). Mapping multilingual lexical semantics for knowledge organization systems. *The Electronic Library*, 30(2), 278-294.
- Chirita, P., Gavriloaie, R., Ghita, S., Nejd, W., & Paiu, R. (2005). Activity based metadata for semantic desktop search. En *The Semantic Web: Research and Applications* (págs. 439-454). Springer.
- Chodorow, M. S. (1985). Extracting semantic hierarchies from a large on-line dictionary. *Proceedings of the 23rd annual meeting on Association for Computational Linguistics* (págs. 299-304). Chicago, IL, USA: Association for Computational Linguistics.
- Chrysafiady, K., & Virvou, M. (2013). A knowledge representation approach using fuzzy cognitive maps for better navigation support in an adaptive learning system. *SpringerPlus*, 2(81). doi:doi:10.1186/2193-1801-2-81
- Classification Research Group. (1955). The need for a faceted classification as the basis of all methods of information retrieval. *Library Association Record*, 57(7), 262-268.
- Cohen, B., & Murphy, G. L. (1984). Models of Concepts*. *Cognitive Science*, 8(1), 27-58.
- Colomo-Palacios, R., García-Crespo, A., Soto-Acosta, P., Ruano-Mayoral, M., & Jiménez-López, D. (2010).). A case analysis of semantic technologies for R&D intermediation information management. *International Journal of Information Management*, 30(5), 465-469.
- Corcho, O., Fernández-López, M., & Gómez-Pérez, A. (2003). Methodologies, tools and languages for building ontologies. Where is their meeting point? *Data & knowledge engineering*, 46(1), 41-64.
- Cross, V. (2004). Fuzzy semantic distance measures between ontological concepts. *Fuzzy Information*, 2, 635-640.
- Cumming, M. (2003). Tomatoes are not the only fruit: a rough guide to taxonomies, thesauri, ontologies and the like. *Records Management*, 113, 11-12.
- Daconta, M. C., Obrst, L. J., & Smith, K. T. (2003). *The Semantic Web. A guide to the future of XML, Web Services, and Knowledge Management*. . Indianapolis: Wiley.
- Dahlberg, I. (2006). Knowledge organization: a new science? *Knowledge organization*, 33, 11-19.
- Davis, F., Bagozzi, R., & Warshaw, R. (1989). User Acceptance of Computer Technology: A Comparison of Two Theoretical Models. *Management Science*, 35(8), 982-1003.
- Davis, R., Shrobe, H., & Szolovits, P. (1993). What is a knowledge representation? . *AI magazine*, 14(1).
- De Nicola, A., & Missikoff, M. (2016). A lightweight methodology for rapid ontology engineering. *Communications of the ACM*, 59(3), 79-86.
- Draper, G., Livnat, Y., & Riesenfeld, R. (2009). A Survey of Radial Methods for Information Visualization. *IEEE Transactions on Visualization and Computer Graphics*, 15(5), 759-776.

- Dupret, G., & Piwowarski, B. (2006). Principal components for automatic term hierarchy building. En *String Processing and Information Retrieval* (págs. 37-48). Berlin: Springer.
- Duval, E., Forte, E., Cardinaels, K., Verhoeven, B., Van Durm, R., Hendrikx, K., & Warkentyne, K. (2001). The Ariadne knowledge pool system. *Communications of the ACM*, 44(5), 72-78.
- El-Beltagy, S. R. (2006). KP-Miner: A simple system for effective keyphrase extraction. *Proceedings of the Innovations in Information Technology*, (págs. 1-5). Dubai, Dubai.
- El-Beltagy, S. R., & Rafea, A. (2009). KP-Miner: A keyphrase extraction system for English and Arabic documents. *Information Systems*, 34(1), 132-144.
- El-Haj, M., Balkan, L., Barbalet, S., Bell, L., & Shepherdson, J. (2013). An experiment in automatic indexing using the HASSET thesaurus. *Proceedings of the Computer Science and Electronic Engineering Conference (CEEC)*, (págs. 13-18). Colchester, Inglaterra.
- Fernández-López, M., Gómez-Pérez, A., & Juristo, N. (1997). Methontology: from ontological art towards ontological engineering. *Proceedings of the Ontological Engineering AAAI-97 Spring Symposium Series*. Stanford, MO, USA.
- Fleiss, J. L., & Cohen, J. (1973). The equivalence of weighted kappa and the intraclass correlation coefficient as measures of reliability.”. *Educational and Psychological Measurement*, 33, 613–619.
- Fotzo, H. N., & Gallinari, P. (2004). Learning generalization/specialization relations between concepts: application for automatically building thematic document hierarchies. *Proceedings of the 7th Conference on Computer-Assisted Information Retrieval*, (págs. 143-155). Avignon, Francia.
- Frank, E., Paynter, G. W., Witten, I. H., Gutwin, C., & Nevill-Manning, C. G. (1999). Domain-specific keyphrase extraction. *Proceedings of 16th International Joint Conference on Artificial Intelligence* (págs. 668-673). Estocolmo (Suecia): San Francisco, USA: Morgan Kaufmann Publishers.
- Frantzi, K., Ananiadou, S., & Mima, H. (2000). Automatic recognition of multiword terms: the C-value/NC-valuemethod. *International Journal on Digital Libraries*, 3(2), 115-130.
- Fuhr, N., Tsakonas, G., Aalberg, T., Agosti, M., Hansen, P., Kapidakis, S., & Micsik, A. (2007). Evaluation of digital libraries. *International Journal on Digital Libraries*, 8(1), 21-38.
- Gamon, M., Yano, T., X, S., Apacible, J., & Pantel, P. (2013). *Understanding document aboutness step one: Identifying salient entities*. Technical Report MSR-TR-2013-73, Microsoft Research.
- Gaona García, P. A., Martín-Moncunill, D., Sánchez-Alonso, S., & García, A. F. (2014). A usability study of taxonomy visualisation user interfaces in digital repositories. *Online Information Review*, 38(2), 284-304.
- Gaona-García, P. A. (2014). *Visualización de esquemas de representación de conocimiento para el acceso a recursos en repositorios digitales*. Alcalá de Henares, España: Tesis Doctoral Universidad de Alcalá.
- Gaona-García, P. A., Martín-Moncunill, D., Gordillo-Orjuela, K. E., & González-Crespo, R. (2016). Navigation and Visualization of Knowledge Organization Systems using Virtual Reality Glasses. *IEEE Latin America Transactions*, 14(6), 2915-2920.

- Gaona-García, P. A., Sánchez-Alonso, S., & Montenegro, C. E. (2014). Visualization of information: a proposal to improve the search and access to digital resources in repositories. *Ingeniería e Investigación*, 34(1), 83-89.
- Gaona-García, P. A., Stoitsis, G., Sanchez-Alonso, S., & Biniari, K. (2016). An exploratory study of user perception in visual search interfaces based on SKOS. *Knowledge Organization*, 43(4), 217-238.
- Gaona-García, P., Martín-Moncunill, D., & Montenegro-Marin, C. (2017). Trends and challenges of visual search interfaces in digital libraries and repositories. *The Electronic Library*, 35(1).
- García-Barriocanal, E., & Sicilia-Urbán, M. A. (2003). User Interface Tactics in Ontology-Based Information Seeking. *Psychology e-journal*, 1(3), 243-256.
- Gašević, D., & Hatala, M. (2006). Ontology mappings to improve learning resource search. *British Journal of Educational Technology*, 37(3), 375-389.
- Gašević, D., Djurić, D., & Devedzic, V. (2009). *Model driven engineering and ontology development*. Springer-Verlag Berlin Heidelberg. .
- Gleicher, M., Albers, D., Walker, R., Jusufi, I., Hansen, C. D., & Roberts, J. C. (2011). Visual comparison for information visualization. *Information Visualization*, 10(4), 289-309.
- Gómez-Pérez, A., Fernández-López, M., & Corcho, O. (2004). *Ontological engineering: with examples from the areas of knowledge management, e-commerce and the Semantic Web*. Londres: Springer.
- Graham, M., & Kennedy, J. (2010). A survey of multiple tree visualisation. *Information Visualization*, 9(4), 235-252.
- Gruber, T. R. (1993). A translation approach to portable ontologies. *Knowledge Acquisition*, 5(2), 199-220.
- Guarino, N., & Poli, R. (1995). Formal ontology, conceptual analysis and knowledge representation. *International Journal of Human Computer Studies*, 43(5), 625-640.
- Hanisch, F., & Straßer, W. (2005). How to include visuals and interactivities in an educational computer graphics repository. *Computers & Graphics*, 29(2), 237-243.
- Hartson, H. R., Shivakumar, P., & Pérez-Quinones, M. A. (2004). Usability inspection of digital libraries: a case study. *International Journal on Digital Libraries*, 4(2), 108-123.
- Harzing, A. W., & Van der Wal, R. (2007). Google Scholar: the democratization of citation analysis. *Ethics in Science and Environmental Politics*, 8(1), 61-73.
- Hearst, M. (2009). *Search user interfaces*. New York, NY, USA: Cambridge University Press. .
- Herman, I., Melancon, G., & Marshall, M. S. (2000). Graph visualization and navigation in information visualization: a survey. *IEEE Transactions on Visualization and Computer Graphics*, 6(1), 24-33.
- Hitchcock, S., Woukeu, A., Brody, T., Carr, L., Hall, W., & Harnad, S. (2003). *Evaluating Citebase, an open access Web-based citation-ranked search and impact discovery service. Monograph (Technical Report) Southampton*. Southampton: University of Southampton. Obtenido de <https://eprints.soton.ac.uk/258204/>
- Hjørland, B. (1998). Information retrieval, text composition, and semantics. *Knowledge Organization*, 25(1), 16-31.

- Hjørland, B. (2001). Towards a theory of aboutness, subject, topicality, theme, domain, field, content and relevance. *Journal of the American Society for Information Science and Technology*, 52(9), 774-778.
- Hjørland, B. (2002). Domain analysis in information science; eleven approaches-traditional as well as innovative. *Journal of documentation*, 58(4), 422-462.
- Hjørland, B. (2011). The importance of theories of knowledge: indexing and information retrieval as an example. *Journal of the American Society for Information Science and Technology*, 62(1), 72-77.
- Hjørland, B., & Albrechtsen, H. (1995). Toward a new horizon in information science: domain-analysis. *Journal of the Association for Information Science and Technology*, 400-425.
- Hjørland, B., & Kylliesbech Nielsen, L. (2001). Subject Access Points in Electronic Retrieval. *Annual Review of Information Science and technology*, 35, 3-51.
- Hodge, G. (2000). *Systems of Knowledge Organization for Digital libraries. Beyond traditional authority files*. Washington D.C., WA, USA: the Council on Library and Information Resources.
- Hu, B., Dasmahapatra, S., & Lewis, P. (2007). Semantic metrics. *International Journal of Metadata, Semantics and Ontologies*, 2(4), 242-258.
- Hughes, B., & Kamat, A. (2005). A metadata search engine for digital language archives. *D&Lib Magazine*, 11(2).
- Iqbal, R., Murad, M. A., Mustapha, A., & Sharef, N. M. (2013). An analysis of ontology engineering methodologies: A literature review. *Research journal of applied sciences, engineering and technology*, 6(16), 2993-3000.
- Isfandyari-Moghaddam, A., & Bayat, B. (2008). Digital libraries in the mirror of the literature: issues and considerations. . *The Electronic Library*, 26(6), 844-862.
- Jarrar, M., & Meersman, R. (2009). Ontology engineering—the DOGMA approach. En E. J. Chang, & K. Sycara, *Advances in Web Semantics I* (págs. 7-34). Springer-Verlag Berlin Heidelberg.
- Jonathan, L., Jinjuan, H. F., & Harry, H. (2009). *Research Methods In Human-Computer Interaction*. John Wiley & Sons.
- Jones, S., & Paynter, G. W. (2001). Human evaluation of Kea, an automatic keyphrasing system. *Proceedings of the 1st ACM/IEEE-CS joint conference on Digital libraries*, (págs. 148-156). Roanoke, VA, USA.
- Jovanović, J., Gašević, D., Knight, C., & Richards, G. (2007). Ontologies for effective use of context in e-learning settings. *Educational Technology & Society*, 10(3), 47-59.
- Julien, C. A., Tirilly, P., Dinneen, J. D., & Guastavino, C. (2013). Reducing subject tree browsing complexity. *Journal of the American Society for Information Science and Technology*, 64(11), 2201-2223.
- Kashyap, V., Morales, A., & Hongsermeier, T. (2006). On implementing clinical decision support: achieving scalability and maintainability by combining business rules and ontologies. *Association of Moving Image Archivist 2006 Symposium*, (págs. 414-419). Anchorage, AK, EE.UU.
- Kennan, M., & Wilson, C. (2006). Institutional repositories: review and an information systems perspective. *Library management*, 27(4), 236-248.

- Khan, S., Baig, A. R., & Shahzad, W. (2014). A novel ant colony optimization based single path hierarchical classification algorithm for predicting gene ontology. *Applied Soft Computing*, 34-49.
- Khoo, M., Kusunoki, D., & MacDonald, C. (2012). Finding Problems: When Digital Library Users Act as Usability Evaluators. *45th Hawaii International Conference on System Science (HICSS)* (págs. 1615-1624). Honolulu, HI, USA: IEEE.
- Kida, M., Tonoike, M., Utsuro, T., & Sato, S. (2007). Domain classification of technical terms using the web. *Systems and Computers in Japan*, 38(14), 11-19.
- Kim, S., & Cavedon, L. (2011). Classifying Domain-Specific Terms Using a Dictionary. *Australasian Language Technology Association Workshop*, (págs. 57-65). Camberra, Australia.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J., & Linkman, S. (2009). Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51(1), 7-15.
- Kotis, K., & Vouros, G. A. (2006). Human-centered ontology engineering: The HCOME methodology. *Knowledge and Information Systems*, 10(1), 109-131.
- Krug, S. (2005). *Don't make me think: A common sense approach to web usability*. . Pearson Education India.
- Kumar, N., & Srinathan, K. (2008). Automatic keyphrase extraction from scientific documents using N-gram filtration technique. *Proceedings of the ACM Symposium on Document Engineering*, (págs. 199-208). Sao Paulo, Brasil.
- Lancaster, F. W. (1972). *Vocabulary control for information retrieval*. Washington D.C.: Journal of the American Society for Information Science.
- Lei, M. (2008). Knowledge organization systems (KOS). *Knowledge Organization*, 35(2), 160-182.
- Lenat, D., & Guha, R. V. (1990). Cyc: A midterm report. *AI magazine*, 11(3), 32.
- Lim, V. M., Wong, S. F., & Lim, T. M. (2013). Automatic keyphrase extraction techniques: A review. *Proceedings of the IEEE Symposium on Computers & Informatics*, (págs. 196-200). Langkawi, Malaysia.
- Lord, P., Macdonald, A., Lyon, L., & Giarretta, D. (2004). From data deluge to data curation. *Proceedings of the UK e-science All Hands meeting*, (págs. 371-357). Nottingham, UK.
- Lytras, M., & Sicilia, M. (2007). Where is the value in metadata? 2(4). *International Journal of Metadata, Semantics and Ontologies*, 2(4), 235-241.
- Madhu, G., Govardhan, A., & Rajinikanth, T. K. (2011). Intelligent Semantic Web Search Engines: A Brief Survey. *International Journal of Web & Semantic Technology*, 2(1).
- Mai, J. (2004). Classification in context: Relativity, reality, and representation. *Knowledge Organization*, 31(1), 39-48.
- Mai, J. (2005). Analysis in indexing: Document and domain centered approaches. *Information Processing and Management*, 41, 599-611.
- Mai, J. (2006). Contextual analysis for the design of controlled vocabularies. *Bulletin of the American Society for Information Science and Technology*, 33(1).
- Martín-Moncunill, D. (2012). *Estudio de usabilidad de motores de búsqueda en la web de organic.edunet.eu*. Information Engineering Research Unit, Departamento de Ciencias de la Computación. Alcalá de Henares: Universidad de Alcalá.

- Martin-Moncunill, D., Gaona-García, P. A., Garcia-Barriocanal, E., & Sanchez-Alonso, S. (2015). Selection and Use of Search Mechanisms in Learning Object Repositories: the Case of Organic. Edunet. *IEEE Revista Iberoamericana de Tecnologías del Aprendizaje*, 11.
- Martin-Moncunill, D., Gaona-García, P. A., Gordillo-Orjuela, K. E., & Montenegro-Marin, C. (2016). User Experiences in Virtual Reality Environments Navigation Based on Simple Knowledge Organization Systems. *Applied Computer Sciences in Engineering. WEA 2016. Communications in Computer and Information Science*. 657, págs. 39-49. Cham: Springer.
- Martín-Moncunill, D., García-Barriocanal, E., Sicilia, M. A., & Sanchez-Alonso, S. (2015). Evaluating the Practical Applicability of Thesaurus-Based Keyphrase Extraction in the Agricultural Domain: Insights from the VOA3R Project. *Knowledge Organization*, 42(2), 76-89.
- Martín-Moncunill, D., García-Barriocanal, E., Sicilia, M. A., & Stracke, C. M. (2017). Evaluating the Concept Specialization Distance from an end-user perspective: the case of AGROVOC. *Working Paper, Accepted with major revision at Online Information Review*.
- Martín-Moncunill, D., García-Barriocanal, E., Sicilia, M. A., & Stracke, C. M. (en prensa). Evaluating the Concept Specialization Distance from an end-user perspective: the case of AGROVOC. *Online Information Review*.
- Martín-Moncunill, D., Sánchez-Alonso, S., Gaona-García, P. A., & Marianos, N. (2013). Applying visualization techniques to develop interfaces for educational repositories: the case of Organic. Lingua and VOA3R. *Proceedings of the Learning Innovations and Quality: The Future of Digital Resources*, (págs. 60-67). Roma, Italia.
- Martín-Moncunill, D., Sicilia, M. A., García-Barriocanal, E., & Sanchez-Alonso, S. (2015). Evaluating the degree of domain specificity of terms in large terminologies: the case of AGROVOC. *Online Information Review*, 39(3), 326-345.
- Mascardi, V., Cordì, V., & Rosso, P. (2007). *Technical Report DISI-TR-06-2*. Università degli Studi di Genova: Dipartimento di Informatica e Scienze dell'Informazione. Obtenido de <http://www.disi.unige.it/person/MascardiV/Download/DISI-TR-06-21.pdf>
- Mascardi, V., Locoro, A., & Rosso, P. (2009). Automatic ontology matching via upper ontologies: A systematic evaluation. *IEEE Transactions on Knowledge and Engineering*, 22(5), 609-623.
- McCarthy, J., & Hayes, P. (1968). Some philosophical problems from the standpoint of artificial intelligence. En B. L. Webber, & N. J. Nilsson, *Readings in Artificial Intelligence* (págs. 463-502). Los Altos, CA, USA: Morgan Kaufmann.
- Medelyan, O., & Witten, I. H. (2005). Thesaurus-based index term extraction for agricultural documents. *Proceedings of the EFITA/WCCA Joint Congress on IT in Agriculture*, (págs. 1122-1129). Vila Real, Portugal.
- Medelyan, O., Frank, E., & Witten, I. H. (2009). Human-competitive tagging using automatic keyphrase extraction. *Proceedings of the International Conference of Empirical Methods in Natural Language Processing*, (págs. 1318 - 1327). Singapur, Singapur.
- Medelyan, O., Milne, D., Legg, C., & Witten, I. H. (2009). Mining meaning from Wikipedia. *International Journal of Human-Computer Studies*, 67(9), 716-754.

- Merčun, T., Žumer, M., & Aalberg, T. (2012). FrbrVis: An information visualization approach to presenting FRBR work families. *International Conference on Theory and Practice of Digital Libraries* (págs. 504-507). Hannover, Alemania: Springer.
- Milne, D., Medelyan, O., & Witten, I. (2006). Mining domain-specific thesauri from Wikipedia: A case study. *Proceedings of the 2006 IEEE/WIC/ACM International Conference on Web Intelligence*, (págs. 442-448). Washington, USA.
- Mizoguchi, R., & Ikeda, M. (1998). Towards ontology engineering. *Japanese Society for Artificial Intelligence*, 9(10), 13.
- Mizoguchi, R., Vanwelkenhuysen, J., & Ikeda, M. (1995). Task ontology for reuse of problem solving knowledge. En N. J. Mars, *Towards Very Large Knowledge Bases: Knowledge Building & Knowledge Sharing 1995* (págs. 46-57). Los Press.
- Morita, T., Tamagawa, S., & Yamaguchi, T. (2014). Constructing a Class Hierarchy with Properties by Refining and Aligning Japanese Wikipedia Ontology and Japanese WordNet. *Proceedings of the 9th International Conference on Ontology Matching-Volume* (págs. 178-179). Aachen, Alemania: ACM.
- Nafria, J. M., Alemany, F. S., & Pérez-Montoro, M. (Febrero de 2017). *glossariumBITri - Interactivo*. Obtenido de <http://glossarium.bitrum.unileon.es/>
- Nash, S. (2005). Learning objects, learning object repositories, and learning theory: Preliminary best practices for online courses. *Interdisciplinary Journal of E-Learning and Learning Objects*, 1(1), 217-228.
- Neches, R., Fikes, R., Finin, T., Gruber, T., Patil, R., Senator, T., & Swartout, W. (1991). Enabling technology for knowledge sharing. *AI magazine*, 12(3), 40.
- Nguyen, D., & Kan, M. Y. (2007). Keyphrase extraction in scientific publications. En *Asian Digital Libraries. Looking Back 10 Years and Forging New Frontiers* (págs. 317-326). Springer Berlin - Heidelberg.
- Nielsen, J. (1994). *Usability Engineering*. San Francisco, CA, USA: Morgan Kauffman.
- Nielsen, J. (13 de Mayo de 2001). *Search: Visible and Simple*. Obtenido de Nielsen Norman Group: Evidence-Based User Experience Research, Training, and Consulting: <https://www.nngroup.com/articles/search-visible-and-simple/>
- Nikolas, A., Sotiriou, S., Zervas, P., & Sampson, D. G. (2014). The open discovery space portal: A socially-powered and open federated infrastructure. En *Digital systems for open access to formal and informal learning* (págs. 11-23). Springer International Publishing.
- Niles, I., & Pease, A. (2001). Towards a standard upper ontology. *international conference on Formal Ontology in Information Systems*, (págs. 2-9). Ogunquit, Maine, EE.UU.
- Niles, I., & Pease, A. (2003). Mapping WordNet to the SUMO ontology. *Proceedings of the 2003 International Conference on Information and Knowledge Engineering*, (págs. 23-26). Las Vegas, USA.
- Niles, I., & Terry, A. (2004). The MILO: A General-purpose, Mid-level Ontology. *international conference on information and knowledge engineering*, , (págs. 15-19). Las Vegas, Nevada, EE.UU.
- Norman, G. (2010). Likert scales, levels of measurement and the “laws” of statistics. *Advances in Health Sciences Education*, 15(5), 625-632.
- Norrie, M. C., Reimer, U., Lippuner, P., Rys, M., & Schek, H. J. (1994). Frames, objects and relations: three semantic levels for knowledge-based systems. *Proceedings of the*

- Workshop on Reasoning about Structured Objects: Knowledge Bases meets Databases*, (págs. 1-5).
- Object Management Group (OMG). (2013). *The unified modeling language specification, version 2.5*. Obtenido de <http://www.omg.org/spec/>
- Ochoa, X., & Duval, E. (2009). Automatic evaluation of metadata quality in digital repositories. *International Journal on Digital Libraries*, 10(2-3), 67-91.
- Ochoa, X., Klerkx, J., Vandeputte, B., & Duval, E. (2011). On the use of learning object metadata: The GLOBE experience Towards ubiquitous learning (pp. 271-284): Springer. En *European Conference on Technology Enhanced Learning* (págs. 271-284). Springer Berlin Heidelberg.
- Palavitsinis, N., Ebner, H., Manouselis, N., & Sanchez-Alonso, S. (2011). Using e-learning technologies and standards to make educational content available: the Organic. Edunet approach. *Journal of Information Technology in Agriculture*, 4(1).
- Palavitsinis, N., Manouselis, N., & Sánchez-Alonso, S. (2009). Evaluation of a metadata application profile for learning resources on organic agriculture. *Proceedings of the Research Conference on Metadata and Semantic Research (MTSR 2009)* (págs. 270-281). Springer Berlin Heidelberg.
- Palavitsinis, N., Manouselis, N., & Sánchez-Alonso, S. (2009). Evaluation of a metadata application profile for learning resources on organic agriculture. *Metadata and Semantic Research*, 270-281.
- Palavitsinis, N., Manouselis, N., & Sanchez-Alonso, S. (2014). Metadata quality in digital repositories: Empirical results from the cross-domain transfer of a quality assurance process. *Journal of the Association for Information Science and Technology*, 65(6), 1202-1216.
- Paranjpe, D. (2009). Learning document aboutness from implicit user feedback and document structure. *Proceedings of the 18th ACM conference on Information and knowledge management* (págs. 365-374). Hong Kong, China: ACM.
- Park, C. A., Bello, S. M., Smith, C. L., Hu, Z. L., Munzenmaier, D. H., Nigam, R., . . . Reecy, J. M. (2013). The Vertebrate Trait Ontology: a controlled vocabulary for the annotation of trait data across species. *Journal of Biomedical Semantics*, 4(1), 13.
- Park, J. R. (2009). Metadata quality in digital repositories: A survey of the current state of the art. *Cataloging & classification quarterly*, 47(3-4), 213-228.
- Pinto, H. S., Staab, S., & Tempich, C. (2004). Towards a fine-grained methodology for Distributed, Loosely-controlled and evolving Engineering of oNTologies. *Proceedings of the 16th European Conference on Artificial Intelligence* (págs. 393-397). Valencia, España: IOS Press.
- Pouliquen, B., Steinberger, R., & Ignat, C. (2003). Automatic annotation of multilingual text collections with a conceptual thesaurus. *Proceedings of the workshop in Ontologies and Information Extraction (EUROLAN 2003)*, Bucarest, Rumanía.
- Purday, J. (2009). Think culture: Europeana. eu from concept to construction. *The Electronic Library*, 27(6), 919-937.
- Rius, A., Sicilia, M. A., & Barriocanal, E. G. (2007). Justificación y Descripción del Dominio de Conocimiento de una Ontología para la Formalización y Automatización de Escenarios

- Educativos. *Simposio Pluridisciplinar sobre Diseño y Evaluación de Contenidos Educativos*. Bilbao, España.
- Rubin, J., & Dana, C. (2008). *Handbook of Usability Testing: How to Plan, Design, and Conduct Effective Tests*. Wiley Publishing .
- Ruecker, S., Radzikowska, M., & Sinclair, S. (2011). *Visual interface design for digital cultural heritage: A guide to rich-prospect browsing*. Burlington, VT, USA: Ashgate Publishing.
- Ryutaro, I., Hideaki, T., & Shinichi, H. (2001). Rule Induction for Concept Hierarchy Alignment. *Proceedings of the IJCAI-01 Workshop on Ontology Learning (OL-2001)*, (págs. 26-29). Seattle, WA, USA.
- Salton, G., Wong, A., & Yang, C.-S. (1975). A vector space model for automatic indexing. *Communications of the ACM*, 18(11), 613-620.
- Sánchez, S., & Sicilia, M. (2005). Normative specifications of learning objects and learning processes: towards higher levels of automation in standardized e-learning. *International Journal of Instructional Technology and Distance Learning*, 2(3), 3-12.
- Sanchez-Alonso, S., & Garcia-Barriocanal, E. (2006). Making use of upper ontologies to foster interoperability between SKOS concept schemes. *Online Information Review*, 30(3), 263-277.
- Sánchez-Alonso, S., & Sicilia, M. A. (2009). Using an AGROVOC-based ontology for the description of learning resources on organic agriculture . *Metadata and Semantics*, 481-492.
- Sánchez-Alonso, S., Cáceres, J., Holm, A. S., Lieblein, G., Breland, T. A., Mills, R. A., & Manouselis, N. (2008). Engineering an ontology on organic agriculture and agroecology: the case of the Organic. Edunet project. *Proceedings of the World Conference on Agricultural Information and IT (IAALD AFITA WCCA 2008)*, (págs. 24-27). Tokio, Japón.
- Sánchez-Cuadrado, S., & Morato, J. (11 de 2009). *Glossarium-BITri*. Recuperado el 16 de 07 de 2016, de <http://glossarium.bitrum.unileon.es/Home/organizacion-del-conocimiento-sistema-de>
- Sarasa, A., Canabal, J. M., & Sacristán, J. C. (2009). Agrega: A distributed repository network of standarised learning objects Distributed Computing, Artificial Intelligence, Bioinformatics, Soft Computing, and Ambient Assisted Living. *International Work-Conference on Artificial Neural Networks* (págs. 466-474). Berlin: Springer.
- Sarkar, K., Nasipuri, M., & Ghose, S. (2010). A new approach to keyphrase extraction using neural networks. *International Journal of Computer Science Issues*, 7(2), 16-25.
- Segura, A., García-Barriocanal, E., & Prieto, M. (2011). An empirical analysis of ontology-based query expansion for learning resource searches using MERLOT and the Gene ontology. *Knowledge-Based Systems*, 24(2), 119-133.
- Shneiderman, B., & Plaisant, C. (2006). Strategies for evaluating information visualization tools: multi-dimensional in-depth long-term case studies. *Proceedings of the 2006 AVI workshop on Beyond time and errors: novel evaluation methods for information visualization*. (págs. 1-7). New York, NY, USA: ACM.
- Shneiderman, S. B., & Plaisant, C. (2005). *Designing the user interface 4 th edition*. . Pearson Addison Wesley: USA.
- Sicilia, M. A. (2014). *Handbook of metadata, semantics and ontologies*. . Singapore: World Scientific.

- Sicilia, M. A., García, E., Aedo, I., & Díaz, P. (2003). Representation of Concept Specialization Distance through Resemblance Relations. En *Advances in Soft Computing - Engineering, Design and Manufacturing (Springer Engineering series)* (págs. 173-181). Springer-Verlag.
- Šimek, P., Vaněk, J., Jarolímek, J., Stočes, M., & Vogeltanzová, T. (2013). Using metadata formats and AGROVOC vocabulary for data description in the agrarian sector. *Plant, Soil and Environment*, 59(8), 378-384.
- Sini, M., Lauser, B., Salokhe, G., Keizer, J., & Katz, S. (2008). The AGROVOC concept server: rationale, goals and usage. *Library Review*, 57(3), 200-212.
- Sliwa, J., Gleba, K., Chmiel, W., Szwed, P., & Glowacz, A. (2011). IOEM-ontology engineering methodology for large systems. *International Conference on Computational Collective Intelligence* (págs. 602-611). Springer Berlin - Heidelberg.
- Solodovnik, I. (2013). Il Progetto VOA3R: la costruzione del sociale e semantico per la ricerca e la valutazione dei contenuti nei Repository Open Access. *Bibliotime*, 16(1).
- Stawniak, M. (2012). Searching Content Related by Semantics. En *Space and Time Interactive 3D Multimedia Content* (págs. 223-251). Londres, Inglaterra: Springer.
- Stefaner, M., Wolpers, M., Memmel, M., Duval, E., Specht, M., Börner, & Klemke, R. (2009). *Evaluation of the MACE system*. MACE Joint Deliverable.
- Stevenson, G., Knox, S., Dobson, S., & Nixon, P. (2009). Ontonym: a collection of upper ontologies for developing pervasive systems. *Proceedings of the 1st Workshop on Content, Information and Ontologies*. Heraklion, Grecia.
- Stock, W. (2010). Concepts and semantic relations in information science. *Journal of the American Society of Information Science and Technology*, 61(10), 1951-1969.
- Stoitsis, G., Protonotarios, V., Sanchez-Alonso, S., Martín Moncunill, D., & Kastrantas, K. (2013). Using agricultural learning portals in developing countries: The case of Organic.Edunet. *Proceedings of the EFITA 2013 Conference*. Turín, Italia.
- Stracke, C. M., Sicilia, M. A., Helmstedt, C., Martín-Moncunill, D., & Protonotarios, V. (2013). *Opening Up Access to Scientific Information: Recommendations for Improving Virtual Repositories and Online Communities*. University of Duisburg-Essen, Alemania: University Publisher.
- Stracke, C., Manouselis, N., Sicilia, M. A., Helmstedt, C., Martín-Moncunill, D., & Protonotarios, V. (2013). *Opening Up Access to Scientific Information: Recommendations for Improving Virtual Repositories and Online Communities*. University of Duisburg-Essen, Alemania: University Publisher.
- Tenopir, C. (2003). *Use and users of electronic library resources: An overview and analysis of recent research studies*. Washington D.C., WA, USA: Council on Library and Information Resources.
- The Getty Research Institute. (27 de 03 de 2017). *Art & Architecture Thesaurus*. Obtenido de <http://www.getty.edu/research/tools/vocabularies/aat/index.html>
- Tickoo, O., & Iyer, R. (2017). Knowledge and Ontologies. En *Making Sense of Sensors* (págs. 83-94). Apress.
- Treleaven, J., Battershill, J., Cole, D., Fadelli, C., Freestone, S., Lang, K., & Sarig-Bahat, H. (2015). Simulator sickness incidence and susceptibility during neck motion-controlled virtual reality tasks. *Virtual Reality*, 19(3-4), 267-275.

- Tsakonas, G., & Papatheodorou, C. (2008). Exploring usefulness and usability in the evaluation of open access digital libraries. *Information processing & management*, 44(3), 1234-1250.
- Turney, P. D. (1999). *Learning to extract keyphrases from text*. National Research Council of Canada, Institute for Information Technology, Technical Report ERB-1057. doi:<http://doi.org/10.4224/8913245>
- Turney, P. D. (2000). Learning algorithms for keyphrase extraction. *Information Retrieval*, 2(4), 303-336.
- Valente, A. (2005). Types and Roles of Legal Ontologies. En V. R. Benjamins, P. Casanovas, J. Breuker, & A. Gangemi, *Law and the Semantic Web. Lecture Notes in Computer Science* (págs. 65-76). Springer Berlin-Heilderberg.
- Van Rijsbergen, C. J. (1979). *Information Retrieval*. Londres, Inglaterra: Butterworths.
- Vivaldi, J., & Rodríguez, H. (2010). Finding Domain Terms using Wikipedia. *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC2010)*, (págs. 386-393). Malta.
- Witten, I. H., Paynter, G. W., Frank, E., C, G., Nevill-Manning, & G, C. (1999). KEA practical automatic keyphrase extraction. *Proceedings of the fourth ACM Conference on Digital Libraries*, (págs. 254-255). Berkeley, CA, USA.
- Zeng, M. L., & Chan, L. M. (2004). Trends and issues in establishing interoperability among knowledge organization systems. *Journal of the American Society for Information Science and Technology*, 55(5), 377-395.
- Zhitomirsky-Geffet, M., Erez, E. S., & Judit, B. I. (2017). Toward multiviewpoint ontology construction by collaboration of non-experts and crowdsourcing: The case of the effect of diet on health. . *Journal of the Association for Information Science and Technology*, 68(3), 681-694.

Anexo I: Interfaces existentes en Organic.Edunet

1. Interfaz simple y avanzada para búsqueda por palabras.

The screenshot displays two search interfaces. On the left is a 'Text based search' box with a search input field, dropdowns for 'Resource type' and 'Educational level', and a 'Search' button. On the right is a larger 'Search' section with a 'Search Keyword' input, radio buttons for 'All words' (selected) and 'Exact Phrase', a 'Search' button, and dropdowns for 'Language' (set to 'All'), 'Educational level' (set to 'Any'), 'Order' (set to 'Alphabetical'), and 'Resource type' (set to 'Any').

2. Interfaz de exploración basada en texto

The screenshot shows a 'Browsing' section with two main categories: 'By Language' and 'By Resource type'. Each category lists various options with their respective counts in parentheses.

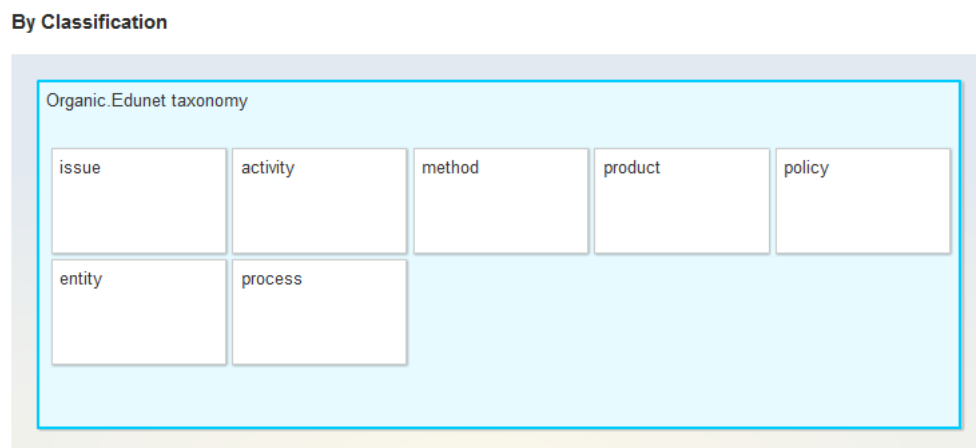
By Language

English (8557)	Estonian (302)	Hungarian (268)	Russian (27)
Norwegian (41)	Spanish (97)	German (1018)	Greek (319)
Romanian (115)			

By Resource type

Application (21)	Assessment (289)	Broadcast (2)	Case study (489)
Course (346)	Demonstration (49)	Drill and practice (27)	Educational game (200)
Enquiry-oriented activity (55)	Experiment (304)	Exploration (882)	Glossary (51)
Guide (832)	Audio (3)	Data (219)	Image (1628)
Model (4)	Text (1041)	Video (101)	Lesson plan (125)
Open activity (26)	Other (425)	Other web resource (121)	Presentation (577)
Project (245)	Reference (1449)	Role play (3)	Simulation (11)
Tool (48)	Weblog (4)	Web page (2238)	Web resource (121)

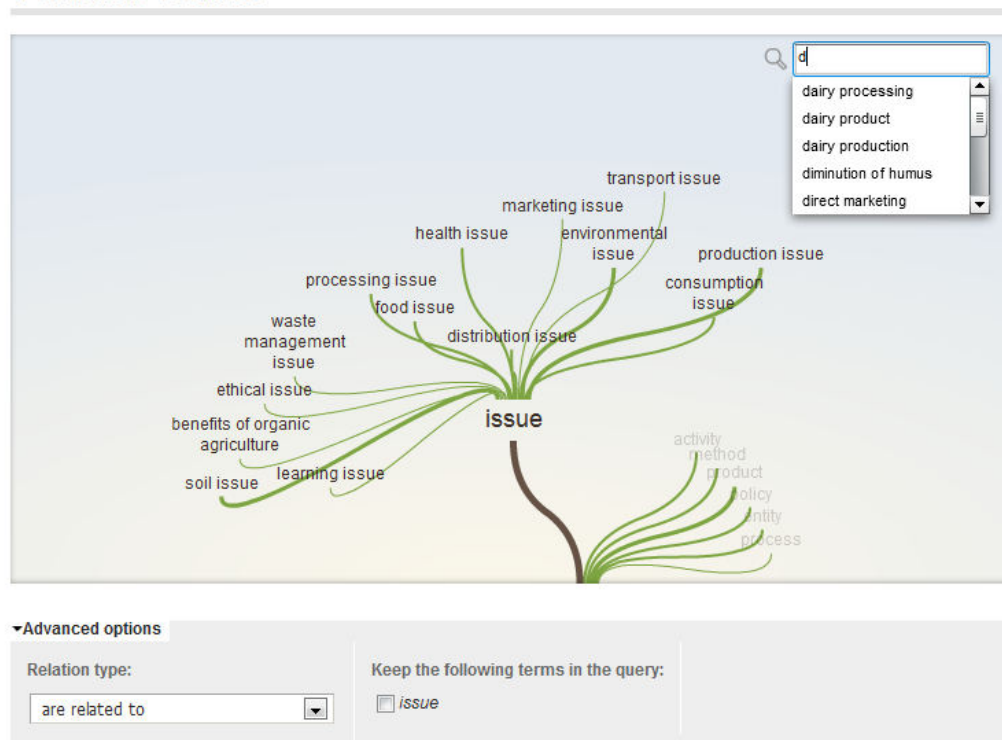
3. Interfaz de exploración visual por categorías





5. Búsqueda semántica

Semantic Search



6. Sugerencias de términos relacionadas con recursos buscados

Resources tagged with: 'agricultura'

Results 1 - 5 of 13

« [Start](#) [Prev](#) **1** [2](#) [3](#) [Next](#) [End](#) »

[Friends of the earth](#)



Children's story which promotes local production and the use of local varieties.



1 votes [View rating history](#) [View reviewing history](#)

Tags: [agricultura](#) [fomentar autóctona enseña infantil región](#) [ADD TAG](#)



7. Búsqueda por escenarios educativos

Educational scenarios

Scenarios and content developed within the Organic.Edunet project

☒ School level

☐ University level

An organic school garden coming into being

[Search](#)

A list of all existing scenarios available through the portal



Handbooks for scenario implementation



School level



University level

The digital resources designed for each scenario

Scenario keyword/hint:

[Search](#)

Anexo II: Guía de aplicación de la metodología para la evaluación de la extracción automática de palabras clave

La evaluación de la efectividad de los algoritmos de extracción automática de palabras clave se ha realizado típicamente comprobando la exactitud del algoritmo para hacer coincidir exactamente las palabras clave proporcionadas por el autor original. Este método resulta muy limitado, puesto que sería necesaria mayor información que reflejara claramente la calidad y la aplicabilidad práctica de estas técnicas en comparación a los resultados que podrían obtener catalogadores humanos y que proporcione respuesta a preguntas como:

- ¿Fueron las palabras clave originalmente seleccionadas por el autor las más apropiadas?
- ¿Humanos expertos en el campo obtendrían mejores resultados que los algoritmos al tratar de coincidir con las palabras clave del autor original?
- ¿Cuál es la calidad relativa de las palabras clave asignadas por el ser humano en comparación con las asignadas de forma automática?

Con tal fin se propone una metodología que amplíe el enfoque tradicional, basado en medidas de precisión y exhaustividad, comparando los resultados obtenidos por el algoritmo con los obtenidos por catalogadores humanos en una evaluación ciega, necesidad que aparece recogida en la revisión de la literatura realizada por El-Haj y otros (2013).

La selección de la muestra sobre la que aplicar la metodología descrita a continuación dependerá de las características de la colección a catalogar y las posibilidades de inversión de tiempo y esfuerzo por parte de aquellos que la vayan a aplicar; de igual forma ocurrirá respecto al número de participantes humanos.

1. Medición de la eficacia del algoritmo en coincidencia exacta

Se partirá de una evaluación clásica de la efectividad del algoritmo empleando el criterio de precisión y exhaustividad (Van Rijsbergen, 1979), comparando los conjuntos de palabras clave obtenidas por el algoritmo con palabras clave asignadas originalmente.

En la evaluación se define como extracción correcta cuando una palabra clave propuesta por el algoritmo coincide exactamente con la palabra clave original; para aquellos KOS que incluyan sinónimos se podría tener en cuenta este hecho como coincidencia exacta.

El número de palabras clave coincidentes se expresará como una proporción de todas las palabras clave extraídas (*Precision* - "P") y del número de palabras asignadas originalmente (manualmente) (*Recall* - "R") para cada documento por separado.

$$P = \frac{\#correct_extracted_keyphrases}{\#all_extracted_keyphrases} \quad R = \frac{\#correct_extracted_keyphrases}{\#manually_assigned_keyphrases}$$

La precisión se calculará dividiendo el número de palabras clave correctamente extraídas por el número total de palabras clave extraídas. La exhaustividad se calculará dividiendo el número de palabras clave extraídas correctamente por el número de palabras asignadas originalmente. Entonces, una combinación equilibrada de los dos se expresa por la medida F, de la siguiente forma:

$$F - measure = \frac{2PR}{P + R}$$

Cuanto más se aproximen a uno estas medidas, mejor será el resultado. En estudios anteriores se han considerado como adecuados valores que se sitúan en la franja de 0.2 a 0.25. Unos valores superiores resultarían estar por encima de lo habitual hasta la fecha.

2. Medición de la eficacia de los catalogadores en coincidencia exacta

La decisión de emplear mecanismos de catalogación automáticos, reemplazaría la catalogación manual por expertos humanos. En este sentido, el siguiente paso sería comparar los resultados obtenidos por catalogadores humanos, al intentar realizar la misma acción que los mecanismos automáticos.

Se solicitará a catalogadores humanos que cataloguen los recursos y se compararán las coincidencias exactas respecto a las palabras clave propuestas por el autor original. Para ello se empleará el mismo criterio de evaluación (precisión y exhaustividad) empleado con los algoritmos automáticos y se compararán los resultados, obteniendo así un indicador sobre la eficacia de los catalogadores humanos respecto a la de los algoritmos automáticos.

3. Evaluación cualitativa por expertos o usuarios

En los pasos anteriores la evaluación quedaba limitada a coincidencias exactas y además toma como patrón válido las palabras clave asignadas originalmente al recurso. Para profundizar en la aplicabilidad práctica de los resultados, se debe tener en cuenta que:

- La elección de las palabras clave está sujeta a un grado de subjetividad. Al aceptar sólo coincidencias exactas el análisis resulta incompleto pudiendo existir términos semánticamente cercanos o con similitud morfológica que también podrían ser útiles desde un punto de vista práctico.
- Asumir que las palabras clave asignadas originalmente, ya sea por el autor o por otro catalogador, son las que mejor representan al recurso, no tiene porqué adecuarse a la realidad.

De este modo, deberá realizarse un análisis con un enfoque más cualitativo en la que podrán intervenir expertos catalogadores humanos o usuarios, atendiendo al contexto y objetivos del estudio.

Se solicitará a los participantes que evalúen la calidad de las palabras clave que se seleccionaron para clasificar los recursos en los pasos anteriores, empleando una escala Likert de 5 puntos donde 1 indique "Representación completamente incorrecta" y 5 "Perfectamente bien representados". Para evitar sesgos en la evaluación, los sujetos en el estudio no deben saber cómo se generaron estas palabras clave (si por otro experto humano, el autor original, automáticamente, por otro usuario, etc).

De esta forma cada evaluador humano evaluará no sólo las palabras clave asignadas automáticamente por el algoritmo, sino también las asignadas manualmente por humanos, impidiendo a los evaluadores conocer cómo se asignaron las palabras clave.

Para la evaluación subjetiva y la tasa de acuerdo entre los sujetos participantes en el experimento se recomienda la utilización del Coeficiente de Correlación Intraclass (ICC) (Fleiss & Cohen, 1973), que representa acuerdos entre dos o más evaluadores o métodos de evaluación sobre un conjunto de sujetos, manejando múltiples observadores y múltiples factores. Este coeficiente se ajusta a los efectos de la escala de mediciones, y autores como Norman (2010) sugieren emplearlo cuando se trabaja con escalas Likert.

Anexo III: Guía de aplicación de la metodología para la evaluación de la especificidad de dominio de un KOS

La creación de KOS suele venir motivada por la necesidad de organizar la información y limitar la complejidad de un dominio concreto, de manera que el KOS pueda ser aplicado para resolver problemas dentro de ese dominio. Sin embargo, los KOS específicos de dominio suelen contener listas de términos genéricos y de fuera del dominio, que pueden limitar su uso. Por otro lado estas listas de términos genéricos y específicos de dentro y de fuera del dominio podrían ser útiles por sí mismos (Kim & Cavedon, 2011) una vez identificados.

Para tal fin se propone la siguiente metodología con la que es posible clasificar los términos de un KOS como propios o no del dominio y establecer una puntuación que define el grado de especificidad global de un KOS o de una parte del mismo.

1. Evaluación del grado de especificidad de los términos por expertos en el dominio

Se solicitará a expertos en el dominio que cataloguen los términos como específicos del dominio o genéricos, entendiendo como término genérico, un término no específico del dominio estudiado. En aquellos casos cuya clasificación no resulte clara, los participantes podrán emplear una escala de 1 a 9, significando 1 "más cercano a ser genérico" y 9 "más cercano a ser específico" mientras que una puntuación de 10 sería considerada como "completamente seguro de ser específico" (S) y 0 "completamente seguro de ser genérico" (G). Adicionalmente se solicitará a los participantes que incluyan observaciones sobre sus decisiones, de cara a profundizar en el análisis cualitativo. Si el participante desconoce un término lo calificará con un 5 y lo marcará en las observaciones. La siguiente tabla refleja un modelo de cuestionario que incluiría toda la información anteriormente indicada:

Tabla 1. Modelo de cuestionario para recoger la información sobre la especificidad de los términos. La segunda fila es un ejemplo.

Término	S/G	No claro	Observaciones
Término del KOS	Indicar S o G	Si no está claro indicar puntuación	<i>"Observaciones relativas al porqué de la elección del participante"</i>
UNCED		5	<i>"Son unas siglas con las que no estoy familiarizado"</i>

2. Clasificación de los términos según su grado de especificidad

Los resultados se recogerán en una sola tabla, asignando los valores numéricos proporcionados por los expertos humanos en el paso anterior para calcular el "grado de especificidad" de cada término. Se calculará la media para cada término y se utilizará este resultado para clasificar cada término como específico o no, de acuerdo con la siguiente escala:

- [0-3]: Genérico / No específico del dominio
- [3-7]: Incierto / No claro
- (7-10]: Específico del dominio

En principio se considera innecesario el uso de criterios de calificación más complejos o de cualquier otro instrumento estadístico debido a que una media dentro de los valores [3-7] representaría uno de los siguientes casos:

- A) Número similar de ocurrencias de valores altos y bajos. Este caso representa una tasa muy baja de acuerdo, dando lugar a un resultado "no claro".
- B) Todos los valores están en el rango [3-7], lo que significa que los expertos no pueden categorizar claramente los términos como específicos o no. Este caso nos lleva también al resultado "no claro".
- C) Una combinación de (a) y (b) que igualmente llevaría al resultado "no claro".

La siguiente tabla muestra una forma adecuada de recoger los resultados de la clasificación de los términos.

Tabla 2. Modelo para recoger los resultados de la clasificación de los términos.

Término / Experto	1	2	3...14	15	16	Media	¿Específico?
Accounting	0	0		0	0	1,25	No
Auditing	0	0		0	0	0,625	No
Depreciation	0	0		0	10	1,5625	No
Seed certification	10	10		10	10	10	Si
Choice of species	10	10		10	0	9,375	Si
Dust extraction	10	10		10	10	6,875	No claro
Grain cleaning	10	10		10	10	10	Si

3. Discusión de los casos inciertos

Con el fin de obtener mayor información sobre el razonamiento de los expertos detrás de la asignación del "grado de especificidad" de un término se recomienda realizar un análisis cualitativo.

Para ello, se deben identificar los casos poco claros y analizar los comentarios de los expertos. Igualmente, se recomienda contactar con algunos de los participantes para obtener información adicional. La tabla 3 muestra cómo se recogieron algunos de estos casos en el experimento *"Evaluating the degree of domain specificity of terms in large terminologies: the case of AGROVOC"* (Martín-Moncunill, Sicilia, García-Barriocanal, & Sanchez-Alonso, 2015) que evaluaba la especificidad de los términos en el dominio de la agricultura del tesoro AGROVOC.

Tabla 2. Ejemplo de análisis cualitativo de los casos poco claros, extraído de (Martín-Moncunill, Sicilia, García-Barriocanal, & Sanchez-Alonso, 2015).

Caso	Motivación
Acronyms such as "UNCED."	<i>They could refer to other institutions, technologies, processes, materials, etc.</i>
Specific terms but from other areas, such as computer science/IT. For instance, "recursive programming."	<i>Some experts got confused; they were told to mark with "G" not only the generic terms, but those not specific to the AGROVOC domains, as depicted in table 3.</i>
Specific terms but from other areas, such as economics. For instance, "environmental accounting."	<i>Same as case "b"</i>
Monitoring and measuring processes such as "gravimetry", or terms related to weight.	<i>They could refer to other areas, even though they are widely used in the domains of agriculture and animal husbandry.</i>
Chemical and physical processes such as "electrodialysis" or "decanting."	<i>Similar to case "e"</i>
Names of common fruits and vegetables. For instance, "apple."	<i>There was no agreement about whether these terms should be considered specific (since they are clearly related to the agricultural and food domains) or generic (since they are widely used terms).</i>
Work-related terms. For instance, "stacking", "pneumatic handling", or "unloading."	<i>They are specific terms related to handling and working. They are widely used in agricultural and forestry domains, but they could be used in many other areas.</i>
Shipping and handling-related terms, such as "packaging" or "wrapping."	<i>Mostly related to food handling in the AGROVOC context but they could refer to other areas.</i>

4. Cálculo del grado de especificidad de dominio de un KOS o de una de sus partes

El primer punto a considerar será si en función de las características particulares del KOS a analizar tendría sentido asignar un peso relativo a su influencia en el grado de especificidad. Por ejemplo, en el experimento al que hace referencia la tabla anterior, se asignaron distintos pesos a cada término de acuerdo a su nivel de profundidad en la jerarquía. En el nivel más profundo (DM) se consideró tener un peso = 1. Así, el peso del término con la profundidad "x" (DX) será:

$$W_x = \frac{D_x}{D_M}$$

De acuerdo a esto, la puntuación máxima de especificidad (S_M - "*Maximum Specificity Score*") será:

$$S_M = \sum_{x=1}^{D_M} T_x * W_x$$

Donde T_x es el número de términos con profundidad = x.

Siguiendo el mismo criterio se calculará la puntuación de especificidad real (S_r - "*Real Specificity Score*"), la puntuación para los casos inciertos (S_u - "*Unclear Specificity Score*") y la puntuación para los casos claros (S_c - "*Clear Specificity Score*"), de la siguiente manera:

$$S_r = \sum_{x=1}^{D_M} T_{sx} * W_x \quad S_u = \sum_{x=1}^{D_M} T_{ux} * W_x \quad S_c = S_M - S_u$$

Donde T_{sx} es el número de términos considerados como específicos con profundidad = x y T_{ux} es el número de términos considerados como inciertos con profundidad = x.

Finalmente se realizará una analogía con el criterio de precisión / exhaustividad propuesto por van Rijsbergen (1979) con el fin de obtener el resultado de especificidad global. El número de términos específicos se expresará como una proporción de todos los términos (*precisión "P"*) y el número de los términos claramente clasificados (*acuerdo "A"*) para cada rama, analizada de forma individual, de la siguiente manera:

$$P = \frac{S_r}{S_M} \qquad A = \frac{S_r}{S_c}$$

El resultado final o grado de especificidad (S - “*Specificity Degree*”) es calculado como una combinación de ambos:

$$S = \frac{2PA}{P + A}$$

Las conclusiones obtenidas deben tener en cuenta que ($P, A, S \leq 1$), o en otras palabras, que el valor máximo que P, A , y S pueden tomar es 1. Si los valores de P y A difieren ampliamente, esto será debido al número de casos poco claros o inciertos. Este grado de incertidumbre (I), puede ser calculado del siguiente modo:

$$I = \frac{S_u}{S_M}$$

Anexo IV: Guía de aplicación de la metodología para la asignación y evaluación de la distancia de especialización entre conceptos

La relación de generalización / especialización ("gen-spec") es una relación entre clasificadores que implica una relación taxonómica y su subsiguiente semántica hereditaria. El entendimiento común de este tipo de relaciones las considera como "todo o nada", asumiendo que la relación siempre es igualmente fuerte entre un clasificador y cualquiera de sus clasificadores relacionados mediante una relación "gen-spec" y también en todos los niveles de la jerarquía. Esto puede ser considerado como una simplificación excesiva del relato psicológico de las relaciones del mundo real (Cohen & Murphy, 2003) (Cross, 2004) (Hu, Dasmahapatra, & Lewis, 2007).

La asignación de un grado de distancia relativa para representar el nivel de similitud entre los pares de clasificadores relacionados podría ser valiosa para fines de búsqueda (Sicilia, García, Aedo, & Díaz, 2003) y se podría aplicar para la representación visual de KOS. Para ello es necesario disponer de algún tipo de metodología en base a la cual asignar este grado de distancia relativa y posteriormente evaluar si realmente esta asignación resultaría de utilidad en el contexto pertinente.

Con tal motivo se propone la siguiente metodología para asignar los valores de distancia de especialización entre conceptos.

1. Asignación de los valores de distancia de especialización entre conceptos

A través de un cuestionario que emplee una escala de puntos, donde el valor más alto represente la distancia más lejana y el valor más bajo la más cercana, se solicitará a un número adecuado de participantes otorgar valores de distancia entre pares de términos. La tabla 1 contiene un extracto de un formulario completado a modo de ejemplo, en un experimento realizado con el tesoro AGROVOC,

Tabla 1. Extracto de un formulario para asignar valores de distancia de especialización entre conceptos. Se está evaluando la distancia entre "Beverages" y el término representado en la siguiente columna.

Términos y jerarquía			Distancia
Foods	Beverages	Alcoholic beverages	2
Foods	Beverages	Coffee	4
Foods	Beverages	Coffee substitutes	4

Una escala de evaluación de 5 puntos debería ser suficiente y adecuada para la mayor parte de los casos, aunque, dependiendo del KOS, puede resultar óptimo emplear otro tipo de escala.

Para apoyar la tarea de asignación del grado de distancia de especialización entre conceptos es recomendable proporcionar a los participantes formularios o interfaces que muestren toda la jerarquía. La tabla 2 muestra un fragmento que contiene 11 términos procedentes del mismo caso que el ejemplo anterior.

Tabla 2. Formulario mostrando la jerarquía completa (extracto de 11 términos)

Distance		Terms (and hierarchy)			
	Foods	Bakery products			
	Foods	Bakery products	Biscuits		
	Foods	Bakery products	Bread		
	Foods	Bakery products	Cakes		
	Foods	Bakery products	Doughs		
	Foods	Bakery products	Puff paste		
	Foods	Bakery products	Batters		
	Foods	Beverages			
	Foods	Beverages	Alcoholic beverages		
	Foods	Beverages	Alcoholic beverages	Beers	
	Foods	Beverages	Alcoholic beverages	Beers	Ales

2. Análisis de los resultados y elaboración del set de datos

Los resultados se recogerán en una tabla y se calcularán los valores de la media y desviación para cada término. La tabla 2 contiene un extracto de resultados a modo de ejemplo en el que participaron 21 sujetos.

Tabla 1. Extracto de una tabla de resultados para cada término.

Términos y jerarquía			1	2	3	17	20	21	Media	Desvia.
foods	Beverages	Alcoholic beverages	3	2	3	3	3	3	2,9048	0,3008
foods	Beverages	Coffee	4	4	4	4	4	4	3,9048	0,3008
foods	Beverages	Coffee substitutes	5	4	4	5	5	4	4,4286	0,5071

Igualmente se recogerán los resultados para cada nivel jerárquico y para toda la selección de términos, tal como se muestra en la tabla 3.

Tabla 3. Resultados globales de un set de datos de distancia de especialización entre conceptos.

Nivel	#Terminos	Distancia Media	Desviación Media
1 --> 2	25	3,820952381	0,443057077
2 --> 3	47	2,80952381	0,402373908
3 --> 4	22	1,413419913	0,393588839
4 --> 5	15	1,038095238	0,095618289
5 --> 6	4	1	0
TOTALS	113	2,654135338	0,353922064

Con el fin de validar los resultados se podrán emplear mecanismos estadísticos que evalúen la tasa de acuerdo. Una vez analizados y validados los resultados se podrá crear el set de datos, originando un KOS que incluya información en cuanto a la distancia de especialización entre sus conceptos.

3. Uso de los resultados para la detección de inconsistencias en el KOS

Dado que para la aplicación de esta metodología, los participantes analizarán los conceptos de un KOS con una visión global de su relación jerárquica, esta podrá resultar útil para:

- Identificar inconsistencias en el KOS.
- Optimización de las relaciones jerárquicas.
- Detectar otro tipo de errores, desde ortográficos a otros más complejos.

Se recomienda a aquellos que deseen aplicar esta metodología, aprovechen la oportunidad para obtener información adicional que facilite la optimización del KOS.

4. Verificación de la utilidad de la información

La información sobre la distancia de especialización entre conceptos extraída en base a los pasos anteriores puede ser aplicada en distintos casos, como la representación de un KOS en una interfaz visual que sitúe los términos en base a la distancia calculada.

Existen multitud de técnicas procedentes del ámbito de la usabilidad que pueden aplicarse para comprobar el resultado del uso de la información de la distancia de especialización entre

conceptos en el KOS objeto de estudio. Entre estas, una prueba ampliamente recomendable sería realizar ensayos A / B comparando la opción que hace uso de esta información con una análoga que no la emplee.